

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 22, 2006, 22:57:33 ; Search time 56 Seconds

(without alignments)  
2291.668 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909

Sequence: 1 MAEAGIRGWLMLALRLAQLA.....GSIKAGAIINFLPNNGROP 1332

Scoring table: BLOSUM62

Searched: 445344 seqs, 96347055 residues

Total number of hits satisfying chosen parameters: 445344

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*  
1: /cgn2\_6/prodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/prodata/1/paa/US11\_NEW\_COMB.pep.\*  
8: /cgn2\_6/prodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6909	100.0	1332	6 US-11-301-094-6	Sequence 6, Appl
2	6872.5	99.5	1359	6 US-11-270-796-22	Sequence 22, Appl
3	5407	78.3	1333	6 US-11-270-796-3	Sequence 3, Appl
4	2402.5	34.8	1278	1 PCT-US06-05584-822	Sequence 822, App
5	2402.5	34.8	1278	6 US-11-191-274A-129	Sequence 129, App
6	2402.5	34.8	1278	6 US-11-191-274A-130	Sequence 130, App
7	1341.5	19.4	1383	6 US-11-301-094-2	Sequence 2, Appl
8	1046	15.1	1274	6 US-11-301-094-4	Sequence 4, Appl
9	782.5	11.3	419	6 US-10-953-429-21065	Sequence 21065, A
10	588.5	8.5	967	6 US-10-461-673-16739	Sequence 16739, A
11	583.5	8.4	1182	6 US-11-332-764-2	Sequence 2, Appl
12	583.5	8.4	1182	6 US-11-337-244-149	Sequence 149, App
13	482	7.0	891	6 US-60-772-265-1197	Sequence 1197, App
14	455	6.6	1358	8 US-60-772-265-241	Sequence 241, App
15	437	6.3	831	7 US-11-360-355-133805	Sequence 133805, A
16	349	5.1	783	6 US-10-461-673-16730	Sequence 16730, A
17	325	4.7	204	7 US-11-360-355-141351	Sequence 141351, A
18	261	3.8	575	7 US-11-360-355-120785	Sequence 120785, A
19	252	3.6	505	6 US-10-461-673-16871	Sequence 16871, A
20	229.5	3.3	465	7 US-11-360-355-141367	Sequence 141367, A
21	225	3.3	542	6 US-11-214-063A-1670	Sequence 1670, Ap
22	222.5	3.2	422	7 US-11-360-355-141365	Sequence 141365, A
23	219	3.2	1137	6 US-10-461-673-16889	Sequence 16889, A
24	219	3.2	1330	6 US-10-461-673-10541	Sequence 10541, A
25	190	2.8	126	7 US-11-360-355-141352	Sequence 141352, A

26	152.5	2.2	632	6 US-11-293-697-3548	Sequence 3548, Ap
27	141	2.0	164	7 US-11-360-355-127069	Sequence 127069, A
28	139.5	2.0	484	7 US-11-360-355-120169	Sequence 120169, A
29	135.5	2.0	252	7 US-11-360-355-140206	Sequence 140206, A
30	123.5	1.8	171	7 US-11-360-355-149589	Sequence 149589, A
31	121.5	1.8	1043	6 US-10-536-606-20	Sequence 20, Appl
32	120.5	1.7	2426	6 US-11-203-806A-11	Sequence 11, Appl
33	118.5	1.7	768	6 US-11-214-063A-2044	Sequence 2044, Ap
34	118.5	1.7	788	6 US-11-214-063A-1692	Sequence 1692, Ap
35	117.5	1.7	619	8 US-60-732-162-1828	Sequence 1828, Ap
36	116.5	1.7	619	6 US-11-312-958-46	Sequence 46, Appl
37	116.5	1.7	619	6 US-10-461-673-12167	Sequence 12167, A
38	114	1.7	488	6 US-60-752-355-45206	Sequence 45206, A
39	114	1.7	985	6 US-11-293-697-2874	Sequence 2874, Ap
40	112.5	1.6	697	6 US-10-703-799B-226	Sequence 226, App
41	111	1.6	201	7 US-11-045-004-2037	Sequence 152528, A
42	110.5	1.6	494	6 US-11-360-355-152528	Sequence 2027, Ap
43	109.5	1.6	451	6 US-11-045-004-2398	Sequence 2398, Ap
44	109.5	1.6	3979	8 US-60-752-355-33716	Sequence 33716, A
45	108	1.6	474	8 US-60-752-355-39891	Sequence 39891, A

## ALIGNMENTS

RESULT 1  
US-11-301-094-6  
; Sequence 6, Application US/11301094  
; GENERAL INFORMATION:  
; APPLICANT: Levitan, Diane J  
; TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS  
; FILE REFERENCE: JB06242US01  
; CURRENT APPLICATION NUMBER: US/11/301,094  
; PRIOR FILING DATE: 2005-12-12  
; PRIOR FILING DATE: 2004-12-15  
; NUMBER OF SEQ NOS: 24  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-301-094-6

Query Match 100.0%; Score 6909; DB 6; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAEAGIRGWLMLALRLAQLAQBEPYTIHQGYCAFYDECGKPELSGIMTLNSVCLSN	60
DB	1	MAEAGIRGWLMLALRLAQLAQBEPYTIHQGYCAFYDECGKPELSGIMTLNSVCLSN	60
QY	61	TPARKITGDHLILLOKICRPLTYGENTQACSAKQVLSLASITKALITRCPACSDNF	120
DB	61	TPARKITGDHLILLOKICRPLTYGENTQACSAKQVLSLASITKALITRCPACSDNF	120
QY	121	VNIHCNTPSPNOSTLINTRVNQLGAGOLPAVVAEAFQHSFAROSYDSCSRVPPA	180
DB	121	VNIHCNTPSPNOSTLINTRVNQLGAGOLPAVVAEAFQHSFAROSYDSCSRVPPA	180
QY	181	ATLAVGTCGVSSALCNARWLNFOGDTGNGIAPLDTFHLLEPGQAVSGIOPINIEGV	240
DB	181	ATLAVGTCGVSSALCNARWLNFOGDTGNGIAPLDTFHLLEPGQAVSGIOPINIEGV	240
QY	241	ARCNEOGDDVATCSQDCAASCPAIAFPQALDSTYLLGMPGSLVLIILCSVFAVVTI	300
DB	241	ARCNEOGDDVATCSQDCAASCPAIAFPQALDSTYLLGMPGSLVLIILCSVFAVVTI	300
QY	301	LIVGFVAPARDSKRVDPKGTSLSDKLSFSHTLLGQFQCGMTWVSWPLTIVLVSV	360
DB	301	LIVGFVAPARDSKRVDPKGTSLSDKLSFSHTLLGQFQCGMTWVSWPLTIVLVSV	360

QY 361 IIVVLAAGLVTELTTPVELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420  
DB 361 IIVVLAAGLVTELTTPVELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420  
QY 421 RYDSLLGKPNFSGILDLDLLELELOERLRLHLOWSPBEAORNISLQDICYAPLNPNT 480  
DB 421 RYDSLLGKPNFSGILDLDLLELELOERLRLHLOWSPBEAORNISLQDICYAPLNPNT 480  
QY 481 SLYDCINSILOYPQNNRTLLLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKOGTALAL 540  
DB 481 SLYDCINSILOYPQNNRTLLLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKOGTALAL 540  
QY 541 SCMAADYGAVPFPFLAIGYKGYKDYSEAEALIMTFSLNYPADGPRLAOKLMEAEFLBEM 600  
DB 541 SCMAADYGAVPFPFLAIGYKGYKDYSEAEALIMTFSLNYPADGPRLAOKLMEAEFLBEM 600  
QY 601 RAFORRMAGMFQVTPFAERSLEDEINRTTAEDLPFAISYIYFLYISLALGSYSMSRY 660  
DB 601 RAFORRMAGMFQVTPFAERSLEDEINRTTAEDLPFAISYIYFLYISLALGSYSMSRY 660  
QY 661 MVDSKATLGLGVAVVLGAVMAAMGFSTYLGIRSSLVILQVVPFLVLSVADNIFIFVLE 720  
DB 661 MVDSKATLGLGVAVVLGAVMAAMGFSTYLGIRSSLVILQVVPFLVLSVADNIFIFVLE 720  
QY 721 YQRLPRRGEPRREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSGLAV 780  
DB 721 YQRLPRRGEPRREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKROBASRLDVCCVCKPELPPPGGGBELLGFPQKAYAPPL 840  
DB 781 ILDFLLQMSAFVALLSLDSKROBASRLDVCCVCKPELPPPGGGBELLGFPQKAYAPPL 840  
QY 841 LHMTRGVVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVAGV 900  
DB 841 LHMTRGVVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVAGV 900  
QY 901 YVVTLLGNFSSBAGMNAICSSAGCNSFSTQKIQYATEPPEOSYIATPSSVWDFFIM 960  
DB 901 YVVTLLGNFSSBAGMNAICSSAGCNSFSTQKIQYATEPPEOSYIATPSSVWDFFIM 960  
QY 961 LPPSSCCRLYISGPNKDFCPSVNSLNCIKNCMSITMGSVRPSVQFHKYLPWFLNDRP 1020  
DB 961 LPPSSCCRLYISGPNKDFCPSVNSLNCIKNCMSITMGSVRPSVQFHKYLPWFLNDRP 1020  
QY 1021 NIKCPKGLAAYSTSVNLTSQGVLA SRFMAVHKPLKNSQDYTEALRAARELANITADL 1080  
DB 1021 NIKCPKGLAAYSTSVNLTSQGVLA SRFMAVHKPLKNSQDYTEALRAARELANITADL 1080  
QY 1081 RYVPGTDPAFEVFPYITTVFYEQYLTILPEGLFMLSCLVPFPFVAVSCULLGLDLSGIL 1140  
DB 1081 RYVPGTDPAFEVFPYITTVFYEQYLTILPEGLFMLSCLVPFPFVAVSCULLGLDLSGIL 1140  
QY 1141 NLLSTIMILVDYTGFMALWDISYNAYSILNLVSAVMSVEPVSHITRSRPAISTKPTWLB 1200  
DB 1141 NLLSTIMILVDYTGFMALWDISYNAYSILNLVSAVMSVEPVSHITRSRPAISTKPTWLB 1200  
QY 1201 AKEBATISMSGSAVFAGVAMTNLPGLIVLGLAKAQLIQIFFEFLNLLITLLGLHLGVFLPV 1260  
DB 1201 AKEBATISMSGSAVFAGVAMTNLPGLIVLGLAKAQLIQIFFEFLNLLITLLGLHLGVFLPV 1260  
QY 1261 ILSYGPDPVNPALALEQKRAEBAVAAVMVASCNHPSTRVSTADNIVNHSFEGSIGAGA 1320  
DB 1261 ILSYGPDPVNPALALEQKRAEBAVAAVMVASCNHPSTRVSTADNIVNHSFEGSIGAGA 1320  
QY 1321 ISNPLPNNGROF 1332  
DB 1321 ISNPLPNNGROF 1332

RESULT 2  
US-11-270-796-22  
; Sequence 22, Application US/11270796  
; GENERAL INFORMATION:

; APPLICANT: Dong, Jianli  
; TITLE OF INVENTION: RESTORATION OF CANCER BY SIMULTANEOUS INHIBITION OF BRAF AND  
; FILE REFERENCE: 02420/1201581-US1  
; CURRENT APPLICATION NUMBER: US/11/270,796  
; PRIOR FILING DATE: 2005-11-08  
; PRIOR APPLICATION NUMBER: 60/592,592  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patencin version 3.3  
; SEQ ID NO 22  
; LENGTH: 1359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-270-796-22

Query Match 99.5%; Score 6872.5; DB 6; Length 1359;  
Best Local Similarity 97.9%; Pred No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY 1 MAEAGLRGMLWALLRLAQBEPYTTIHQPGYCAEYDEGKNPELSIMTLSNVCSLSN 60  
DB 1 MAEAGLRGMLWALLRLAQBEPYTTIHQPGYCAEYDEGKNPELSIMTLSNVCSLSN 60  
QY 61 TPARKITGDHLILLOKICPRLYTGPNTQACCSAKQVLSIABSLSTTKALTFCPCASDNF 120  
DB 61 TPARKITGDHLILLOKICPRLYTGPNTQACCSAKQVLSIABSLSTTKALTFCPCASDNF 120  
QY 121 VNLHCNHTCSPNQSLFINVTRVAQAGQLPAVVAYEAFQHSFAEOSTDCSRVRVPA 180  
DB 121 VNLHCNHTCSPNQSLFINVTRVAQAGQLPAVVAYEAFQHSFAEOSTDCSRVRVPA 180  
QY 181 ATLAVTMCQGVYSALCNAORMLNFQCDTGNGLABLDITFHLBEGQAVGSGIQPLNEGV 240  
DB 181 ATLAVTMCQGVYSALCNAORMLNFQCDTGNGLABLDITFHLBEGQAVGSGIQPLNEGV 240  
QY 241 ANCNESQGDVATCSGCCDCAASCPAIARPQALDSTFYLGOMGSLVLIILCSVAVVTI 300  
DB 241 ANCNESQGDVATCSGCCDCAASCPAIARPQALDSTFYLGOMGSLVLIILCSVAVVTI 300  
QY 301 LLVGRVVPAPARKSMQVDPKGTSLSDKLSFSTHTLLGOFPQGMGTWASWPLTILVLSV 360  
DB 301 LLVGRVVPAPARKSMQVDPKGTSLSDKLSFSTHTLLGOFPQGMGTWASWPLTILVLSV 360  
QY 361 IIVVLAAGLVTELTTPVELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420  
DB 361 IIVVLAAGLVTELTTPVELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420  
QY 421 RYDSLLGKPNFSGILDLDLLELELOERLRLHLOWSPBEAORNISLQDICYAPLNPNT 480  
DB 421 RYDSLLGKPNFSGILDLDLLELELOERLRLHLOWSPBEAORNISLQDICYAPLNPNT 480  
QY 481 SLYDCINSILOYPQNNRTLLLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKOGTALAL 540  
DB 481 SLYDCINSILOYPQNNRTLLLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKOGTALAL 540  
QY 541 SCMAADYGAVPFPFLAIGYKGYKDYSEAEALIMTFSLNYPADGPRLAOKLMEAEFLBEM 600  
DB 541 SCMAADYGAVPFPFLAIGYKGYKDYSEAEALIMTFSLNYPADGPRLAOKLMEAEFLBEM 600  
QY 601 RAFORRMAGMFQVTPFAERSLEDEINRTTAEDLPFAISYIYFLYISLALGSYSMSRY 660  
DB 601 RAFORRMAGMFQVTPFAERSLEDEINRTTAEDLPFAISYIYFLYISLALGSYSMSRY 660  
QY 661 MVDSKATLGLGVAVVLGAVMAAMGFSTYLGIRSSLVILQVVPFLVLSVADNIFIFVLE 720  
DB 661 MVDSKATLGLGVAVVLGAVMAAMGFSTYLGIRSSLVILQVVPFLVLSVADNIFIFVLE 720  
QY 721 YQRLPRRGEPRREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSGLAV 780  
DB 721 YQRLPRRGEPRREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKROBASRLDVCCVCKPELPPPGGGBELLGFPQKAYAPPL 840

Db 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCKVQDELPPPGQEGGLLGFPOKAYAPFL 840  
Qy 841 LHMITRGVLLPLALFGVSLYSMCHISYVGLDOELALPKDSYLLDYFLFLNRFEEVGA 900  
Db 841 LHMITRGVLLPLALFGVSLYSMCHISYVGLDOELALPKDSYLLDYFLFLNRFEEVGA 900  
Qy 901 YFVTTLGYNFSSBAGNNAICSSAGCNNSFTQKIQYATEPPEQSYLAIPASSVWDDP 960  
Db 901 YFVTTLGYNFSSBAGNNAICSSAGCNNSFTQKIQYATEPPEQSYLAIPASSVWDDP 960  
Qy 961 LTPSSCCRLYISGPNNDKCPSTVNSLNCCKMSITMGSVRPSVQPHKYLFWFLNDR 1020  
Db 961 LTPSSCCRLYISGPNNDKCPSTVNSLNCCKMSITMGSVRPSVQPHKYLFWFLNDR 1020  
Qy 1021 NIKCPFGGLAAVSTSVNLSDDQVL-----ASRFMAVH 1053  
Db 1021 NIKCPFGGLAAVSTSVNLSDDQVL-----ASRFMAVH 1053  
Qy 1054 KPLKNSQDYTEALRAARELANITADLRKVPGTDPAEVEPYTTTNVFEQYLLTIDPEGL 1113  
Db 1054 KPLKNSQDYTEALRAARELANITADLRKVPGTDPAEVEPYTTTNVFEQYLLTIDPEGL 1113  
Qy 1114 FMLSCLVPTFAVSCILLGLDLSGLNLSTIMLVDTYGFMAAMDISTNANSLNLYS 1173  
Db 1114 FMLSCLVPTFAVSCILLGLDLSGLNLSTIMLVDTYGFMAAMDISTNANSLNLYS 1173  
Qy 1174 AVGMSVEFVSHITRSPALSTKPTWLERAKATISMGSAVAGVAMTNLPGILLGLAKQ 1233  
Db 1201 AVGMSVEFVSHITRSPALSTKPTWLERAKATISMGSAVAGVAMTNLPGILLGLAKQ 1260  
Qy 1234 LIQIFPFRNLTLTLGLHGLVFLPVLISYVGPDVNPALALEQKAEBAVAAVMAVASC 1293  
Db 1261 LIQIFPFRNLTLTLGLHGLVFLPVLISYVGPDVNPALALEQKAEBAVAAVMAVASC 1320  
Qy 1294 NHPRSVSTADNIVYVNSFEGSIKGAALSNFLPNNGROF 1332  
Db 1321 NHPRSVSTADNIVYVNSFEGSIKGAALSNFLPNNGROF 1359  
RESULT 3  
US-11-270-796-3  
; Sequence 3, Application US/11270796  
; GENERAL INFORMATION:  
; APPLICANT: Dong, Jiani  
; TITLE OF INVENTION: TREATMENT OF CANCER BY SIMULTANEOUS INHIBITION OF BRAF AND  
; FILE REFERENCE: 02420/1201581-US1  
; CURRENT APPLICATION NUMBER: US/11/270,796  
; PRIOR FILING DATE: 2005-11-08  
; PRIOR APPLICATION NUMBER: 60/592,592  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 3  
; LENGTH: 1333  
; TYPE: PRF  
; ORGANISM: Mus musculus  
US-11-270-796-3  
Query Match 78.3%; Score 5407; DB 6; Length 1333;  
Best local similarity 77.2%; Pred. No. 0;  
Matches 1088; Conservative 126; Mismatches 174; Indels 4; Gaps 4;

Db 123 IHCHNTCSPOSQSLFIVNTRVAVQAGQLPAVVAVEAFYQHSFAEQSYDSCSRVAVPAAT 182  
Qy 183 LAVGTCGYGYSALCAQOBNLNFQDGTGNGLAPLDTTFPHLLEBQAVSGSIOPLNGBVAA 242  
Db 183 LAVGTCGYGYSALCAQOBNLNFQDGTGNGLAPLDTTFPHLLEBQAVSGSIOPLNGBVAA 242  
Qy 243 CNEGSGDDVATCSGQDCAASCAPAIARPQALDSTFYLGQMGSLVLIIILCSFAVAVTTIL 302  
Db 243 CNEGSGEDSAACSCQDCAASCAPVIPPALRPSFYGMRPMGLALIIIFPAVFLSVVL 302  
Qy 303 YGRVAPPAADKSMVDPKKGTSLSDLSTFSTHTLLQFOFGQGTWVAWPLTLVLVSVR 362  
Db 303 YVLRVAVSNRNKNKTAGSQAPMLPRRRRSPHVLGRFPESWGTWRAVWPLVLAISFV 362  
Qy 363 VVLAAGLVFTBLTTPDVEILMSAPNSQARSEKAFHDOHGGPPRTNOVLTIANRBSRY 422  
Db 363 VILMSVGLFTIELTTPDVEILMSAPNSQARSEKAFHDEHGGPPRTNOVLTIANRBSRY 422  
Qy 423 DSIILGPKNFSGILLDLLELELELOERLRLHQLQWSPEAQRNLSLQDI CYAPLPNDTSL 482  
Db 423 DSIILGPKNFSGILLDLLELELELOERLRLHQLQWSPEAQRNLSLQDI CYAPLPNDTSL 482  
Qy 483 YDCCNSLLQYFQNNFTLLLTANQTLMGQTSQVQKQHPFLYCANAPLPFKOCTALALSC 542  
Db 483 TDCCVNSLLQYFQNNHTLLLTANQTLMGQTSLVQKQHPFLYCANAPLPFKOCTALALSC 542  
Qy 543 MDVYGPVPPFLAIGGKQDYSBAFALIMTFSINNYPPGDDPLAOKIMEBAFLEBMA 602  
Db 543 IADYGPVPPFLAVGSIQDYSBAFALIMTFSINNYPPGDDPLAOKIMEBAFLEBMA 602  
Qy 603 FQRMAGMFQVFTAESLEDEINRTAEDLPFATSYIVTFLYISLALGSYSWSRWV 662  
Db 603 FQRMAGMFQVFTAESLEDEINRTAEDLPFATSYIVTFLYISLALGSYSWSRWV 662  
Qy 663 DSQATYGLGAVAVVLCVAVNANGFBSYLGIRSSLVTLQVVPFLVSVGADNITIFVLEYQ 722  
Db 663 DSQATYGLGAVAVVLCVAVNANGFBSYLGIRSSLVTLQVVPFLVSVGADNITIFVLEYQ 722  
Qy 723 RLPRRGPBEVHIGBALGVAPSMILCSLEALICFELGALPMPAVRPFALNSGLAVIL 782  
Db 723 RLPRRGPBEVHIGBALGVAPSMILCSLEALICFELGALPMPAVRPFALNSGLAVIL 782  
Qy 783 DFLQMSAFVALLSLDSKROEASRLDVCCCKVQDELPPPGQEGGLLGFPOKAYAPFLH 842  
Db 783 DFLQMSAFVALLSLDSKROEASRLDVCCCKVQDELPPPGQEGGLLGFPOKAYAPFLH 842  
Qy 843 WITRGVLLPLALFGVSLYSMCHISYVGLDOELALPKDSYLLDYFLFLNRFEEVGA 902  
Db 843 WITRGVLLPLALFGVSLYSMCHISYVGLDOELALPKDSYLLDYFLFLNRFEEVGA 902  
Qy 903 VTTLGYNFSSBAGNNAICSSAGCNNSFTQKIQYATEPPEQSYLAIPASSVWDDP 962  
Db 903 VTTLGYNFSSBAGNNAICSSAGCNNSFTQKIQYATEPPEQSYLAIPASSVWDDP 962  
Qy 963 P-SSCCRLYISGPNNDKCPSTVNSLNCCKMSITMGSVRPSVQPHKYLFWFLNDRPN 1021  
Db 963 P-SSCCRLYISGPNNDKCPSTVNSLNCCKMSITMGSVRPSVQPHKYLFWFLNDRPN 1021  
Qy 1021 IKCPFGGLAAVSTSVNLSDDQVLRAAFMAVYHKKPLKNSQDYTEALRAARELANITADLR 1081  
Db 1021 IKCPFGGLAAVSTSVNLSDDQVLRAAFMAVYHKKPLKNSQDYTEALRAARELANITADLR 1081  
Qy 1082 KVPGTDPAEVEPYTTTNVFEQYLLTIDPEGLFMLSCLVPTFAVSCILLGLDLSGLNL 1141  
Db 1082 KVPGTDPAEVEPYTTTNVFEQYLLTIDPEGLFMLSCLVPTFAVSCILLGLDLSGLNL 1141  
Qy 1142 LLSIWMILVDTYGFMAAMDISTNANSLNLYSVAVSGSVFVSHITRSPALSTKPTWLER 1201  
Db 1142 LLSIWMILVDTYGFMAAMDISTNANSLNLYSVAVSGSVFVSHITRSPALSTKPTWLER 1201  
Qy 1202 KEATISMGSAVAGVAMTNLPGILLGLAKQDLIQIFPFRNLTLTLGLHGLVFLPV 1261  
Db 1202 KEATISMGSAVAGVAMTNLPGILLGLAKQDLIQIFPFRNLTLTLGLHGLVFLPV 1261

Db 1203 KQATIFMGSAGVAMTNFPGILLIGFAQAQLIQIFFFRNLNLTLLGLHGVLPV 1262  
Qy 1262 LSYGPDVNPALALEQKRAEAVAAVMAVASCNHPERSVSTADNIYNHSEES-IKGAGA 1320  
Db 1263 LSYLGPDVQAQALVLEEKATEA-AMVSESPSCQYPPPADANTSVDYNYGFNBEFPEINA 1321  
Qy 1321 ISNPLNNROF 1332  
Db 1322 ASSSLPKSDQK 1333

## RESULT 4

PCT-US06-05584-822  
Sequence 822, Application PC/TUS0605584  
GENERAL INFORMATION:  
APPLICANT: DANA-FARBER CANCER INSTITUTE, INC.  
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION,  
FILE REFERENCE: DFS-064.25(125992-6425)  
CURRENT APPLICATION NUMBER: PCT/US06/05584  
CURRENT FILING DATE: 2006-03-02  
PRIOR APPLICATION NUMBER: 60/690,064  
PRIOR FILING DATE: 2005-06-13  
PRIOR APPLICATION NUMBER: 60/654,227  
PRIOR FILING DATE: 2005-02-17  
NUMBER OF SEQ ID NOS: 848  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 822  
LENGTH: 1278  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US06-05584-822

Query Match 34.8%; Score 2402.5; DB 1; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 1.1e-185;  
Matches 522; Conservative 227; Mismatches 450; Indels 103; Gaps 25;

Qy 7 RGMILMALILRLAQSEPTTTHQPCYARYDCG-----KNPELSSGLMTLSVSCLS 59  
Db 4 RGLALGLLTLCLCPAQVFSQ-----SCWYGEGLAVGDKRYNCEYSG----- 46  
Qy 60 NTPARKITGDHLLLOKICPRLYTGPNTQACSAQVLSLEASISTTKALTRCPACSDN 119  
Db 47 --PYKLPDQGYDLVQELCPGFFFG-NVSLCCDVRLQTLKONLQLOPLQFSLRCSCEFN 103  
Qy 120 FVNLCHNTCSBNQSLFINVTR-----VAQLGAGQLPAAVAAFYQHSFAEQSYDSCSRV 175  
Db 104 LNLFCCLTSPRQSGFLNVTATEDVVDVTNQTKNVKELOQYVGQSFANMAYNACRPV 163  
Qy 176 RVPAAATLAVGTMCQVYGSALCNAQWLNFGQDGTGNGLAPLDT-----PHLEPGQA 228  
Db 164 EAPSSNDKALGLCGKDADA-CNATWIEYMFNKGQAPFTITPVFSDFPVH----- 215  
Qy 229 VSGGIQPLNEGVARCNESGDDVATCSGDDCAASGALAPQ-----ALDSTFVIG 279  
Db 216 ---GMEPMNATKGCDESDVETAPCSGDDCIVGPKRPPPPAPMTLLGLDAMYVM 272  
Qy 280 QMPGSLVLIILCSYFAVNTILL-----VGRVAPARDKSKVDPKKGTSLS 326  
Db 273 WITWAFLLVFGAFPAVWCYKRKYVSYTPIDSIANSV-NASKGE-----ASGC 324  
Qy 327 DKLSTSTHTLLGQFQSGKGTWASWPLTLVLSVIVVALAAGLVFTELTTDPVELWASP 386  
Db 325 DVSAAFEBCGLRRLFRMGSCFVRNPGCVIFPSLVETITACSSGLVFRVTTNPVDLMSAP 384  
Qy 387 NSQASEKAFHQQHQPFRTRNOVILLANRSRYRDLGLPK-NFSGIILDLDELTEL 445  
Db 385 SSOARLEKGFPOHGFPRTEQLIRAPLTDKHTIOPYPSGADVFGPPLDIQILHQL 444  
Qy 446 ELQERLRHLQVWSPEAQRNISLQDICVAPLNPNTSLDYCCINSLLQYFQNNRTLLTLTA 505  
Db 445 DLQIAIEN--ITASYNETVTIQDICLAPLSPYNT--NCTILSVANTYQNSHSLVDHKK 499

Qy 506 NOTLMGQTSQVMDKHFLYCANAPLTFKDGKALALSCMADVGAPVFPFLAIGYKQDYS 565  
Db 500 GDDF---VYADYHTHTFLYCVRAPASLNDTSLHDPCLGTGCGVFPVLVAGYDDQNTN 556  
Qy 566 EAEALIMTFSLNINYPADGPRLAQAKLEAEALFEEMRAFORMAAGMFOVTFTAERLDEI 625  
Db 557 NATALVITFPNNYNDTEKQRAQAEKEFINPKYKNT--PULTISFTERSIEDEL 613  
Qy 626 NRTAEDLPITATSYIVFLYISLALSGYSKSRVNDSKATLGLGVAUVLGAAMAAG 685  
Db 614 NRESDVFTVVISYAIMFLYISLALGHIKSCRLLVDSKSLGAGILYLVSSVACSIG 673  
Qy 686 FFSYGISSLVILLOVPELVSVAGDNIFIVLEYORLPRRGPBREHGRALGRVAP 745  
Db 674 VFSYIGLPTLIVLEIPPLVAVGVNDIFILVQAVORDERQGTLLQOQLGRVLGEVAP 733  
Qy 746 SWLCSLSEALCFPLGALTMPAVATFPLTSGLAVIDLFLQMSAFVALLSLDSKROBAS 805  
Db 734 SMFLSSFEBEYAFGLALSNVPAVHTFSLFAGLANFIDFLQITCFVGLLGLDIRQEN 793  
Qy 806 RIDVCCVYKPOELPPPGQ-GEGLLIGFQKAYAPFLHMTTRGVLLFLALFGVLSYSM 864  
Db 794 RLDIFCCVRGAEDGTSVQASESCLEFRPKNSYSPLLKDMRPVIAIFVGLSFSIAVL 853  
Qy 865 CHISVGLDQELALPQDSYLLDYFLFLNRYFVGAVYFVTLGYNFSSSEAGNNALCSSAG 924  
Db 854 NKVDIGLQDSLSMPDSTYWDVFKSISQYLHAGPVPYVLEBGHDYTSKQGMVCGGNG 913  
Qy 925 CNNSFTOKIOYATEFPQSYLAIPASSWVDPLDMLTP-SGCCRLYISGNKDKFCPSY 983  
Db 914 CNNDLSLVQOIFMAQALDNYTRIIGFAPSSMIDYDFPMVAFQSSCCGV---DNITDQFNAS 970  
Qy 984 VNSLNLCKNCSIT-MGSVRPSVEQFKYLLPFLNDRPNIKCPKGLAAYSTSVN--LTS 1040  
Db 971 VDPACV-CBPLTPEBQKQRCQGDPMRFLPMFLSDNPNPCGKGHAAYSAVAILGH 1029  
Qy 1041 DQVLAASFMAVHKRLKNSODYTEALRAARELANITADLRKVPETDAFEVFPYTTNV 1100  
Db 1030 GTRVATYPMYHTVLTQTSADPIDALKKARLASHVT-ETMGINS--AYRVPYSVTV 1086  
Qy 1101 FVEQYLTLPBGLFMLSCLVPTFAVSCLLGLDRLSGSLNLDSIMVLVNDTVGFMALMD 1160  
Db 1087 FVEQYLTIIIDITIFLWGLSLAIFLYTMVLLGCELSAVINCATIAMLVNFMFGMILM 1146  
Qy 1161 ISNVAISLINIVASGVSEFVSHITRSFATSTKPTWLERAKATISGSAVPAVAMTN 1220  
Db 1147 ISLNAVSIVNLVMSGSISEFCSHITRAFTVSMKSGRVERAEALAHMGSSVFSGITLT 1206  
Qy 1221 LRGILVGLAKAQLQIFFRNLNLTLLGLHGVLPVILSYVGPVNPV 1272  
Db 1207 FQGIIVLAFAKSQIFQIFFRMYLAMVLLGATHGLIFLPLVLLSYGPSVNKA 1258

## RESULT 5

US-11-191-274A-129  
Sequence 129, Application US/11191274A  
GENERAL INFORMATION:  
APPLICANT: Appleira Corporation  
APPLICANT: Bruno DOMO  
TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF  
FILE REFERENCE: C1001536PROV  
CURRENT APPLICATION NUMBER: US/11/191,274A  
NUMBER OF SEQ ID NOS: 334  
SOFTWARE: FaastSeq for Windows Version 4.0  
SEQ ID NO 129  
LENGTH: 1278  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-191-274A-129  
Query Match 34.8%; Score 2402.5; DB 6; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 1.1e-185;



Matches	522;	Conservative	237;	Mismatches	450;	Indels	103;	Gaps	25;
QY	7	RGMLMALRLAQA	SEPTTTHQPGCAFYDECG-----	KNPELSGSLMTLSNVSCLS	59				
Db	4	RGALGILLLLLC	CPAVFSG-----	SCWVGEGCIAVGDKRYNCERSG-----	46				
QY	60	NTPARKITGDHLL	LOKICPRLYTGPTQACSAKOLVSLBASITTKALLTRCPACSDN	119					
Db	47	--PPKELPKDGYDL	VELCPGPFPG--NVSILCCDVRLOTLKONLOPLQPLSRCPSCFYN	103					
QY	120	PVNLHCHNTCS	PNQSLFINVTR-----	VAOLGAGOLPAVVAEAFYQHSFAEGSYDCSRV	175				
Db	104	LNLFLCELTCSP	QOFLNVTATEDYVDPVTNQTNNVKELOYYVGQSFANAMYNACRDV	163					
QY	176	RVPAAATLAVGT	MCVYGSALCNAORWLNFGDGTGGLAPLDT-----	PHLEBGOA	228				
Db	164	EAPSSNDKALG	ILCGDADA--CNATWIEIMFKMDGQAPFTITTPFSDEPVH-----	215					
QY	229	VSGGIQPLNEG	VARCNESQGDVAVTCSQODCAASCALARPQ-----	ALDSTFYLG	279				
Db	216	---GMEPMNATK	GCDESVDEVTAPCSQCDSIVCPKPPPPAPWTLIGLDAMYIM	272					
QY	280	QMRGSLVLIIL	ICSVAVVTILL-----	VGRVAPARDKSKMVDKKGTSLS	326				
Db	273	WTTMAFLVFG	AFVAVWCYRRRYVSEYTPIDSNIAFSV--NASDKG-----	ASCC	324				
QY	327	DKLSFSTHTLL	CGFQMGKTWVAMWPLTILVSVIPVVALAAGLVTELTTPDEVELMSAP	386					
Db	325	DVVSAAFEGCL	RRLFTRMGSFCVRNPGCVIFFSLVITACSSGLVFRVRYTNVDLMSAP	384					
QY	387	NSQARSEKAF	HDHGFEPFRFNQVILTPNRSSRYDSLLGPK--NFSGILDLLEL	445					
Db	385	SSQARLEKEY	FDOHGFEPFRTEBOLIRAPLTDKHIQYPPSGADVFGPPLDQILHQVL	444					
QY	446	ELQERLRHLQ	WSPRQARNISLQDICYAPLNPNTSLYDCINSLLQYFQNNRTLLLLTA	505					
Db	445	DLQIALEN--	ITASYNETVTLQDICTAPISPYNT--	NCTILSVLNFYNSHVSVDHKK	499				
QY	506	NOTLMQOTQ	VDWKDFLYCANAPLTFKQGTALALSCMADYGAVPFPFLAIGCYKGDYS	565					
Db	500	GDDEF--	YVADYHTFLCYVRAPASLNTSLHDCLGTFGGPVPFWLVLGGYDDQNTN	566					
QY	566	EABALIMTSL	NNYPAGDPRLAQAKLMEAFLEMKRAFORRMAGMFQVFTARSLEDEI	625					
Db	557	NATVALITE	PVNNYVNDTEKLOQAQWKEKFINFVKYKN--	PULTISPTARSIJEDL	613				
QY	626	NRTTADLPI	FATSYVIFYISLAGSYSSMRVWDSKATIGLGVAVVLCAWAAMG	685					
Db	614	NRESDDVFT	VVSVAIMPLYSIALGHKSKRRLLVDSKVSIGIAGILVLSSVACSLG	673					
QY	686	PFYSYLGIR	SSLVLOVVPFLVLSVGADNPIFVLEYQRLPRRGPBPREVHIGBALRVAP	745					
Db	674	VSYISGLP	LTLYEIVPLVLAVGVDNFIILVQYQDERLQGETLDDQGLGVLEGVAP	733					
QY	746	SMILCSLSEAI	CFFLGATPMPARTFALTSGLAVILLQMSAFVALSLDSKROEAS	805					
Db	734	SMFLSFSFE	SVAFFLGLSVMRAVHTFSLFAGLAVFIDLQITCFVSLGLDIIKQEKX	793					
QY	806	RUDVCCCVK	PQELPPGQ--GEGILLGFQKAYAPFLMLHTTRGVALLFLALFGVSLYSM	864					
Db	794	RDLDFCCV	GABGDSVQASECLFRFPFKNSYSPLLKKMMRPIVIAIFGVLSFSIAVL	853					
QY	865	CHISVGLD	DELAKPSYLLDLYFLPLNRYEVEGAPVYFVTTLLGYNSSSEAGMAICSSAG	924					
Db	854	NKVYDILDS	LSMPSDSIMVDYFKSISQYIHAGRPVYFVLEBHDYTSKQGMVCGMG	913					
QY	925	CNNSFTOKI	QVATEPQSYLAIPASSWVDFIDWLTLP--SSCCRLYISGPNKDKCPST	983					
Db	914	CNNDLSLV	QOIFNNAQODNTRIGAFAPSSWDYFDMWKPGQSSCCRV--	DNIIDQCNAS	970				
QY	984	VNSLNLK	KCMGTT--MGSVRSPVEQFHKXIPWFLNDRPNKCPKXGLAANSTSVN--LTS	1040					
Db	971	VVDPACVR--	CRPLTPGKQRPQGGDFWRFLPMFLSDNPNKCGKGGHAAVSSAVNILLGH	1029					

QY	1041	DGOVLASR	MAHYHKLKNSODYTBALPAARELANITADLRKYPGTDPAFEVPPYTTNV	1100
Db	1030	GTRVGATY	FMTYHTVLTQTSADFLDKARLLIANSNT--ETMGJNGS--AYRVPYSVIFY	1086
QY	1101	FYEQVYTL	ILPEGLFMILSLCPFFFAVSCILIGLDSGLLNLISYWLIVDTVGFPMALMD	1160
Db	1087	FYEQVYTL	ITDDTTFNLGVSLGALFVYTWVLGCELSAVIMCATIAVAVLVNMGVMMLKG	1146
QY	1161	ISYNAVSL	INLVASGVSEFVSHITRSPALSTKPTMLERAKBATISMGSAVAGVAMTN	1220
Db	1147	ISLNAVSL	VNLVWMSGISVBFCHITRAFVSKSGRVEABEALAHMSSSVSGIILTK	1206
QY	1221	LPGLVLAGA	KAOILQIFFRNLNLLTLGLHGLVLPVITISYGVDPVNA	1272
Db	1207	FGGIIVLA	FAKSQIFQIFFRMYLAVLLGATGILFLPLVLSYIGSVAKA	1258

## RESULT 6

US-11-191-274A-130  
 ; Sequence 130, Application US/11191274A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Applera Corporation  
 ; APPLICANT: Bruno DOMO  
 ; TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF  
 ; FILE REFERENCE: CL001536PROV  
 ; CURRENT APPLICATION NUMBER: US/11/191,274A  
 ; CURRENT FILING DATE: 2005-07-28  
 ; NUMBER OF SEQ ID NOS: 334  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 130  
 ; LENGTH: 1278  
 ; TYPE: PR1  
 ; ORGANISM: Homo sapiens  
 ; US-11-191-274A-130

Query Match 34.8%; Score 2402.5; DB 6; Length 1278;

Best Local Similarity 39.8%; Pred. No. 1,1e-185;

Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

QY	7	RGMLMALRLAQA	SEPTTTHQPGCAFYDECG-----	KNPELSGSLMTLSNVSCLS	59
Db	4	RGALGILLLLLC	CPAVFSG-----	SCWVGEGCIAVGDKRYNCERSG-----	46
QY	60	NTPARKITGDHLL	LOKICPRLYTGPTQACSAKOLVSLBASITTKALLTRCPACSDN	119	
Db	47	--PPKELPKDGYDL	VELCPGPFPG--NVSILCCDVRLOTLKONLOPLQPLSRCPSCFYN	103	
QY	120	PVNLHCHNTCS	PNQSLFINVTR-----	VAOLGAGOLPAVVAEAFYQHSFAEGSYDCSRV	175
Db	104	LNLFLCELTCSP	QOFLNVTATEDYVDPVTNQTNNVKELOYYVGQSFANAMYNACRDV	163	
QY	176	RVPAAATLAVGT	MCVYGSALCNAORWLNFGDGTGGLAPLDT-----	PHLEBGOA	228
Db	164	EAPSSNDKALG	ILCGDADA--CNATWIEIMFKMDGQAPFTITTPFSDEPVH-----	215	
QY	229	VSGGIQPLNEG	VARCNESQGDVAVTCSQODCAASCALARPQ-----	ALDSTFYLG	279
Db	216	---GMEPMNATK	GCDESVDEVTAPCSQCDSIVCPKPPPPAPWTLIGLDAMYIM	272	
QY	280	QMRGSLVLIIL	ICSVAVVTILL-----	VGRVAPARDKSKMVDKKGTSLS	326
Db	273	WTTMAFLVFG	AFVAVWCYRRRYVSEYTPIDSNIAFSV--NASDKG-----	ASCC	324
QY	327	DKLSFSTHTLL	CGFQMGKTWVAMWPLTILVSVIPVVALAAGLVTELTTPDEVELMSAP	386	
Db	325	DVVSAAFEGCL	RRLFTRMGSFCVRNPGCVIFFSLVITACSSGLVFRVRYTNVDLMSAP	384	
QY	387	NSQARSEKAF	HDHGFEPFRFNQVILTPNRSSRYDSLLGPK--NFSGILDLLEL	445	
Db	385	SSQARLEKEY	FDOHGFEPFRTEBOLIRAPLTDKHIQYPPSGADVFGPPLDQILHQVL	444	
QY	446	ELQERLRHLQ	WSPRQARNISLQDICYAPLNPNTSLYDCINSLLQYFQNNRTLLLLTA	505	

```
Db      445 DQIAIEN--ITASVDNETVTLQDICTLAPLSPYNT--NCTLLSYNTVQNSHSYLDHKK 499
      506 NOTLNGQTSQVDMKHFILYCANAPLTFKDGATLALSCMADYGA PVFPPFIAIGSKYKDY 565
      500 GDDFF---VYADYHNFHLCVAPASLNDTSLHDPCLQTFGGPVPVPMVLVGGYDQNN 556
      566 EAEALIMFSLNNYPAGDRLAOKIMEAFLEEMKAFORRMAGFOVTFMERSLIEDI 625
      557 NATALVITFPVNNVNDTEKLOPAQMEKEFINFVKNYKN---PNLTISFTLERSIEDL 613
      626 NRTTAEDELIFATSVIVIFLYISLALGSYSMSRWVWDSKATLLGCGVAVVVGAWMAAG 685
      614 NRESOSDFTVVISTAINMFLYISLALGHYKSCRLKLVDSKVSIGLAGLITLSSVACSIG 673
      686 PFSYLGISSLVILQVPPFLVLSVGADNIFIFVLEYORLPRRGPBEVHIGRALGRVAP 745
      674 VFSYIGLPPLTLVIEVIFPLVAVGVDNIFILVQAVQRBRIQGETLDQGLRIVEGAVP 733
      746 SKLLCSLSEALCFPLGALTTPMPAVRTFALTSGLAVIDLFLQMSAFVALLSIDSKRQEA 805
      724 SMFLSSFSETVAFPLGALSVMMPAVHFFSLFAGLAVFIDELQITCFVSLGLDIKKQEK 793
      806 RLVCCCKVPOELPPRGQ--GEGLLGFRQKAVAPFLIMHITRGVVLLEFLALRGVSLYM 864
      794 RLIDFCVRGAGDGTSVQASCLPFRFFKNSYSPLLLKDMRPVIAITVGLSFSIAVL 853
      865 CHISVGLQDELALPKDSYLLDYFLFLNRYFEVGA PVPYFTTGLYSSBAGMNAICSSAG 924
      854 NKVDIGLSDLSMPDSDSYVDYFKSISQYLHAGRPYFLYLBEGHDVTSKGNMVCGMG 913
      925 CNNSFFTKIQYATEFPEGSYLAIPASSVVDFFDMLTP--SSCCRLYISGPNKDFCPS 983
      914 CNNDLSLVQOIFENAAQLDNTRIGFAPSSWIDYFDVWVKPQSSCRV---DNITDQCNAS 970
      984 VNSLNCCKMCSIT--MGSVRPVEOECHKLPWFLNDRBNIKCPKCGLAYSTSVN--LTS 1040
      971 VVDPAQVR--CRPLTPBGKORPOGGDWRLPMLFSLNPNPKCGKGAHYSSAVNLLGH 1029
      1041 DQOVLASRFMAHYKPKLNSQDYTEALRARELELANITADLRKVPGTDPAEFVPPYTTNV 1100
      1030 GTRVGATYFMTHYVLTQTSADPFDALKKARLASANT--ETMGINGS--AYRFPYSVFV 1086
      1101 FVEQYVLTIPBGLFMLSCLVPTFAVSCLLGLGLRSGLLNLSIVMILVDTVGFMALWD 1160
      1087 FVEQYVLTIDDTIFNIGVSLGAIFFLVTVLGCGLMSAVIMCATIAMVLVNMFGVMMWG 1146
      1161 ISYNAVSLINIVSACMSVEFVSHITRSPALSTKPTWERAKEATISMGSAVPAVAMTN 1220
      1147 ISLNAVSLVNLVMSCGISVEFCSHITRAFTVSNKSGSRVERAEALAHMGSSVFSGITLTK 1206
      1221 LFGIVLGLAKAQLIQIFFRNLNLTLLGLLHGLVFLVILSYVGPVNPA 1272
      Db      1207 PGGIVVLAPAKSQIFQIFFRMYLNAVILGATHGLIFLVLVLSIGPSYNKA 1258

RESULT 7
US-11-301-094-2
; Sequence 2: Application US/11301094
; GENERAL INFORMATION:
; APPLICANT: Levitan, Diane J
; APPLICANT: Smith, Marsha
; TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS
; FILE REFERENCE: JB06242US01
; CURRENT APPLICATION NUMBER: US/11/301, 094
; CURRENT FILING DATE: 2005-12-12
; PRIOR APPLICATION NUMBER: 60/636,390
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 1363
; TYPE: PRN
```

```
; ORGANISM: Caenorhabditis elegans
US-11-301-094-2
Query Match      19.4%; Score 1341.5; DB 6; Length 1363;
Best Local Similarity 27.5%; Pred. No. 1.5e-99;
Matches 369; Conservative 257; Mismatches 561; Indels 155; Gaps 40;

      76 KICPRLTYGPNTQACCSAKOLVSLASISITKALLTRCPAGSDNVNHLCHNTCPSPNGL 135
      63 BECPHLLTGDN--KLCCTPSQAEGLTKQIAOAHILIGRCPSCPDNPAKIMCEPTSPNOD 121
      136 FINTVRVAGL--GAGQLPAVVAYEAF-----YQHS--FAQSYDSCSRVRPAATLAVG 186
      122 FVISEMKPIEKKEGFTPEYQPAEYVNTVEYRLSTDPAEGMFSCKDYTFBGQALRV- 180
      187 TWCQVYGALCNAOBLNFOGDTNGT--APDITFHLLEPGQAVSGIOP--LNEGVARCN 244
      181 -MC---TSPICLTNMLEIFGTQNDLNIPIHTKFLVDPKITPPSDSTVNNVNTGCD 236
      245 BSQGDVATCSCQDQACASCPAIAFPQALDSTFYLGOM----PGLVLIIILCSFPAVVT 299
      237 KSARVGMWPAQSTSEC--NKEBYANLIDDDGKTSQGTQNVHGIACLNIFVMLAFGLSLAV 294
      300 ILLVGF-----RVAPADSKXVNDPKKGSLSADKLSFTHTLLGQRF 341
      295 LLCVGFVFTSYDEDTYTNLRQTOGSESPRRNRK---RTGAMT-----HNTMENNA 342
      342 QCMGTWVASWPLTIIIVLSVIPVVALAGLVTELTTPDEVLMSAPNSQARSEKAFHQHF 401
      343 RDIGMAGRNPKSHFFIGCAVLIPLCPMITHKESNTNVDMMSRPSRARQGEVFNANF 402
      402 GPFFRTNOVILTAPNRSSIRYDLSLLGPRNSGILDLLELLELLEQLERHLQVWSPBA 461
      403 GRPQRYQOIMLL--SHRFQSGSKLYGP---VFKHDI FEEELPIILNAIKNISITQSDSG 455
      462 QNITLQDICYALNLPDNTSLYDCCINSLLQYFQNNRTLLLTANQT-----L 509
      456 -RTITLDVDCRAMPBG---YDCLIMSPNTNFQGNKEHLDMKSKEETVSGDDDAFDYF 510
      510 MGOATQVDMKHFILYCANAPLTFKDGATLALSCMADYGA PVFPPFIAIGSKYKDYSEABA 569
      511 SEEATTDMMNTMAACIOQPSQK--TKSGLSQMGTYGSPAPRN--VFGKNSHTNQAAANS 567
      570 LIMFSLNNYPAGDRLAOKIMEAFLEEMKAFORRMAGFOVTFMERSLIEDINRT 629
      568 IMMTLLVVO--RTEBEIOKAEIMEKEFLKFCYERKSPKV--FSGMAERSITDEIENDA 624
      630 AEDLPIFATSVIVIFLYISLALGSY-----SSMSRWVWDSKATLLGCGVAVVVGAMAM 684
      625 KQEIYTVVIALAFELIGYVTFSLGRYFCBNQLMS--ILVHSRI CLDMLSYITNLLSPCSW 683
      685 GFYSYLGISSLVILQVPPFLVLSVGADNIFIFVLEY--QRLP---RRRGPBEVHIGRA 739
      684 GIFSMFGIHPVKNALVNGEFTVTLTGVCRTFNVVXYVQQRVSMYMSDQCEP--VGVAV 742
      740 LGRVAPSMLLCSLSEALCFPLGALTTPMPAVRTFALTSGLAVIDLFLQMSAFVALLSIDS 799
      743 MAGTWPAMFSSSLGCAFSFICGFTDLPAIRFTFCYAGLAVLIDVALLCTTIFLAFVNDT 802
      800 KQOASRLDVCCCVKPOELPPRGQEGU---LLG-----FQKAAAPFLHMI 844
      803 QRELNG-----LP--EFFEYQIKDLGAVLIGRQRTDTMTQFHFPOVAPFLMHM 853
      845 TGGVLLLEFLALFGVLSVMCHIISVGLQDELALPKDSYLLDYFLFLNRYFEVGA VVYFT 904
      854 TIIITGIIFASFITTVILSSKISIVGPDQSMAFTEKSIYSHFRILDXFEDVGPVFFV 913
      905 TLGYNFSEBAGMNAICSSAGCNNSFTQKIQYATEFPEGSYLAIPASSVVDFFDMLT-P 963
      914 DDELDMHPRVDQNKCTFPFGCSDTSFGNIMNVAVGHTEQYILSGMYNVIDVLEWISRK 973
      964 SSCRLYISGPKDKDFCPSYVNSL---NCLKNM-----SIMGSV---RPSVEQCH 1009
```

Db 974 SPECYVYVDPN--TFQSTNRKNSALDQACRTQMDPDVANSYPKSIMGHPSIEVEY 1031  
Qy 1010 KTLPMFLNDRPNKCPKGGSLAAYSTSVNLTSQGVLAASFMAVHKL--KNSODYTEAR 1067  
Db 1032 RHLRHLLETPNSECFFGGASPFKQKISFTSRGRLOASQPFTHFKLSJNSSDFTKAD 1091  
Qy 1068 AARELAANTADLRKVPCTDPAFVEVPYTTTNNFEYEQYLTLLPEGLFMLSCLVFPFAVS 1127  
Db 1092 TARMVSRRLEBSI-----DQTAHVAVYAKIRFFPYQYSTIMPILTQQLPIYVGVFGII 1145  
Qy 1128 CULLGLDRSGLLNLLSIYMLVDVYGFPMALMDISYNAVSLNLVSAVMSVEFVSHIR 1187  
Db 1146 CTTGLIDVGAACAVALCQVSNYFHYAFMYIENIPYNALSATNLVMSGILIEFSVNLK 1205  
Qy 1188 SFAITKTPWLBRAKATISMSASVAVGAMTNLPBILVGLAKAQLQIFPRNLTLT 1247  
Db 1206 GVACSLRORAKRABSTVSGISGPIILSGVVYTAGSTWMLSGAHQIITVYFPKLPILTI 1265  
Qy 1248 LGLLHGLVFLPVILSYVGPDPVNPALALBQKAEAAVMAVMSCPNHSRSTADNIVY 1307  
Db 1266 VSSAVHALIILPILAFGSGRGGSGSETSTNDNEQHDACVLS--PTASHISNVEBGI 1323  
Qy 1308 N-----HSPEGSIKGAGAI 1321  
Db 1324 NRPSLIDASHIIDLPLKAGGI 1345  
RESULT 8  
US-11-301-094-4  
Sequence 4, Application US/11301094  
GENERAL INFORMATION:  
APPLICANT: Levitan, Diane J  
APPLICANT: Smith, Marsha  
TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS  
FILE REFERENCE: JB06242US01  
CURRENT APPLICATION NUMBER: US/11/301, 094  
CURRENT FILING DATE: 2005-12-12  
PRIOR APPLICATION NUMBER: 60/636,390  
PRIOR FILING DATE: 2004-12-15  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 4  
LENGTH: 1274  
TYPE: PRF  
ORGANISM: Caenorhabditis elegans  
US-11-301-094-4  
Query Match 15.1%; Score 1046; DB 6; Length 1274;  
Best Local Similarity 25.7%; Pred. No. 1,2e-75;  
Matches 345; Conservative 241; Mismatches 587; Indels 168; Gaps 46;  
Qy 10 LMALLLRAGSEPYTTTHIQPGYCAFYDCGKRPBLSGIMTLSNVSCLSNTPARKITGD 69  
Db 14 VFLFLIHLALCO-----AKCVN--TECDGEEDSNHPCKTNKSTYLPITVTRSLNPT 64  
Qy 70 HLILQICPRPYTGPN--TOACCSAKQVLSLEASLSTTRALLTRCPACSDNPFNLHCNT 128  
Db 65 VWAREPKYSYLVQEBDKQVCTELQKGMTRISNATILIGSCSCFDNFAKLMCOPT 124  
Qy 129 CSPNOSLFINTRVAOLGAGQLPAVVAYEAFYQHSFAEOSYDSCSRVVPAAATLAVGTM 188  
Db 125 CSPDQSKFMKVMET-----TGPKVVVVKMEFKVNRDVEGLYESCRHTWFGANGLALMSL 180  
Qy 189 CGVYSALCNAQRMLNFOGDTG--NGLA---PLDITHLEPGQAVSGIQPLNEGARCN 244  
Db 181 GGRKVS-----FENFYGMGTKNLAOSIPINTEQFSRMKANMIPPTP-----CH 225  
Qy 245 ESQGDVATCSQDCD---AASCPAIPALDSTFIQWPGSLVLIILCSVFAVVTLL 301  
Db 226 KSRGPRVPAAGALDCTPNAHQVLDISKVEHTKTVHNPFR--DFEWLTKICGLA-LTYL 283  
Qy 302 LVGF-----RVAPARDKSKVDPKKGTSLSDKLSFSTHTLLGQFPQG---WGTWVAS 350

Db 284 LVFILLKYSCHRBSAPNGEDGCYVDLQGN-----LEVOEGLCARYANAVIK 330  
Qy 351 WPLTLLVLSVIVVALLAAG-LVFTELTDPVEILMSAPNSQARSEKAF-HDQHGCPFFPTN 408  
Db 331 HPLIFVSLIVIAAACCSNPFHSHLTHSVDOVASADGSTRREKFKFHS--FGPHRIE 388  
Qy 409 QVILTAPNSSRYRYSLLGLPKNFSGILDLDBLLELLELOERLRHLOVMSPEAORNISQ 468  
Db 389 QIFINLPPT-----KSMFMPLPEENFQLVGNIGNLT--ACVGNSSVKD 432  
Qy 469 DICVAPLNDNTSLDCCINSLLQYFQNNRTLLLTANOTLMGQTSQVDMKHFLYCANA 528  
Db 433 DICYKRGKN-----HGCALMSPTNYFQNKMTFENNAPPTIDELFPDQWHEILKYCIRN 488  
Qy 529 PLTFQDGTALLSCMADYCAPVFPPLATGG-----YKQXQ--YSEBAELIMTFSLNTPAGD 583  
Db 489 PLTV--STYSEMSCFEBSFGPIIDPLLVFGSNEBISGAEVYTAARTIMTVLRG----- 541  
Qy 584 PRLAQKLWEEAFLBEMRAFORRMAGMPQVTFEATERSLEDEINRTTAED--LPIFATSYI 641  
Db 542 PE-DQALIMETAFLNMRSRYEMKXANF---TMTETSVABEIHVAVETDKIVSVICAAV 597  
Qy 642 VIFLYISIALGSY--SSWSRVVDSKATLGLGVAVVILGAVVAVLGAVMAAMGFSSYLGRSSLVIL 699  
Db 598 LHMVITMLGINHMPRESSILSALVHHKLISISAVMISVISVMSIGMFSLFGVHARDNAI 657  
Qy 700 QVVPFLVLSVGADNIFIVLEQYQLPRRGEF---REVN--IGRALGVAPBMLLCIS 753  
Db 658 VVLFVITCTGIRNFVLIIRTFQANGCYGLPISRYEMNHRIISNWRISIPVLNLSL 717  
Qy 754 EALCPFL-GALTP-----WPAVRTPALTSLGAVILDFLLQMSAFVLLSLSKRO----- 802  
Db 718 CSTCLFAGGVLPVYSVMPAVEVPRRHAGLIMDTAAYLVMLLPQYDAREMSGKC 777  
Qy 803 -----EASRLDVCCKVQDELPPPCQGBGLLGFQKAYAPFLMHTGCVLLP 853  
Db 778 EIMPWYELSNESKINKCMADVGNLRSP-----YDWKLAIPALPKKICRWATATP 830  
Qy 854 LALFGVSLYSKCHISVGLDQELALPKDSYLLDFLEFLNRYFEVGADEVYVTTLTGINPSS 913  
Db 831 FVSLIACYCTLCLEFGFNQVAFSETSYLTGKHFQNMENLNIIGPLMFEVGEDVAKHP 890  
Qy 914 AGMNAICSSAGCNBSPFOKIQ---YATFPFQSYLAIPASSVNDPDMITP--SSCCRL 969  
Db 891 KMQNKFCITLGGDDMSGNKIRSLAENY--KGNYLHGDVNIWLSYLOFMHRRGSCCKM 949  
Qy 970 YISGPKDKFC--PSTVNSLNCJKNKWSITMGSVRPSVEQFHKYLPPFLNDRPNKCPKGG 1028  
Db 950 -----DGKOFCDPS--NATHC--SSCSSSVASLITTEYFYRNLHHPLETPEBIOCAHGG 1001  
Qy 1029 LAAVSTSVNLTSQGVLAASFMAVHKP--LKNODYTEALRAARELAANITDLARKVPGT 1086  
Db 1002 MALAKRNLITNKGKIQSAVFTFPFKKLNLSQIQYDAMRFKAYLADIDIRL--EIPGV 1060  
Qy 1087 DPAFEVFPYTTTNNFEYEQYLTLLPEGLFMLSCLVTPAVSCLLGLDRSGLLNLSY 1146  
Db 1061 ----KVVYSTFPFYEQYLTLLSTVYTLVVVLFAVFTTISLFTLVNLAGSLVYTFVLL 1116  
Qy 1147 MLVNDVGFMALMDISYNAVSLNLVSAVMSVEFVSHITRSPASTKTPWLBRAEAT 1206  
Db 1117 SSYILHMEMWYLLGITVNVVSYINMMSLGIAVEFGQMLHGYNSKKPRGRERAPALAV 1176  
Qy 1207 SMGSAVFAVAMTNLPBILV---LGLAKAQLQIFFPRLNLTLTLLGLHGLVFPVLI 1262  
Db 1177 SMGSTLSGI-----FALMITAGCLSFADSRVLIITFCNQLVIGLVCVAGHVYVPTLL 1232  
Qy 1263 SYVGPDPVNPALALBQKAEAA 1283  
Db 1233 AIFGSDFYQNVSESESTDEA 1253  
RESULT 9  
US-10-953-349-21065

```
; Sequence 21065, Application US/10953349
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO: 21065
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Glycine max
; US-10-953-349-21065

Query Match          11.3%; Score 782.5; DB 6; Length 419;
Beet Local Similarity 38.4%; Pred. No. 5.3e-55;
Matches 164; Conservative 88; Mismatches 138; Indels 37; Gaps 10;

QY      863 SMC-HISVGLDDELALPKDSYLLDYFLFLNRYFEVGPARYFTTLGYNFSEA-GMAAIC 920
DB      2  ALCTIEAGLEBQIALPRDSYLGQYFNSISEYLARGPPLFYVK-DYNVLESKHTNOLC 60
QY      921 SGAGCNNSFTOKIQYATEFPQSYIALPASSVVDDEFDWLTPS--SCCRLYISGPNKDK 978
DB      61  SISHCDNSLNLNEISRASLVPSTSYAKPAASMLDFLWISPEAFSCCRKF---TNDG 116
QY      979 FCPSPVNSLNCCKNCSTMGSV-----RPVEOPHKLPLPFLNDPNI 1022
DB      117 YCPDDPQPCCLPDDEGCGGLGGCKDCTTCFPHSDLVNDRPTAQRRELPFLDLPSA 176
QY      1023 KCPKGLAAYSTSVNLT--SDQVLAIRFMAVHKPLKNSQDYTEALRAARELAANITADL 1080
DB      177 DCAKGHGAGNTSVNLNGEGVIAQSEERTYHTPLNRGDVYNAIRARDSSAISSSL 236
QY      1081 RKVGTDPAEVFPYTTINVEYEQYTLIPBGLFMLSCLVTPFAVSCILLGLDRSGLL 1140
DB      237 K-----MDIFPYSVFYIFFEQYLDIKLALINTVALGAIFVV-CLITSSVSSSAI 287
QY      1141 NLSTVILVDTVGFPMALMDISYNAVSLINLVSAVMSSEFVSHITRSPALSTKPTWLER 1200
DB      288 LLLVIMITLIDMGVAALIGQLNNAVSVNLISGIAVEFCVHVHAFVMSLQDR-SGR 346
QY      1201 AKEATISGSAFAGVAMTNLPGLIVLGLAKAQLIQIPEFRLNLTITLLGLHGLVFLPV 1260
DB      347 AKTALCTMGASVFSGLTLKLVGLVLCFSTSEIPVYVYFQWYALVITIGFLHGLVFLPV 406
QY      1261 ILSVGP 1267
DB      407 VLSLFGP 413

RESULT 10
; Sequence 16739, Application US/10461673
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Yang, Yonghong
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Ujwal, Manusha L.
```

```
; APPLICANT: Ma, Yuning
; APPLICANT: Chen, Rui-Hong
; APPLICANT: Ghosh, Malabika
; APPLICANT: Weng, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drenthac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 823
; CURRENT APPLICATION NUMBER: US/10/461,673
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: PCT/US02/29964
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 10/245,014
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 60/323,739
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: PCT/US02/29636
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 10/245,817
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 60/323,349
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: PCT/US02/29001
; PRIOR FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: US 10/243,552
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US02/25485
; PRIOR FILING DATE: 2002-08-09
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 17116
; SOFTWARE: pc_fl_genes Version 6.0
; SEQ ID NO 16739
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-461-673-16739

Query Match          8.5%; Score 588.5; DB 6; Length 967;
Beet Local Similarity 21.3%; Pred. No. 1e-38;
Matches 216; Conservative 192; Mismatches 35; Indels 24; Gaps 37;

QY      334 HT-----LLGFQFGMGTVASWPLTILVSLVVALAAGLVF--TELTPDYELWSAP 386
DB      115 HTDCEGLLSRTFQQLGQVGAHPWIFLLAPMLTALAGTGLYLPKDEEBDEBHHYPIV 174
QY      387 NSQARSEKAFHDQHGPFRTNQVILTPNRS-----KYDSDLGPKNFSGIL 436
DB      175 GSPAKAERRFVQGH---FTTNDSYRFSASRSTEFANFVSLVSYSDSLDPATPAEVS 230
QY      437 DLDLLELELELOERLRLQVWSPKARNSLQDIC--YAPINPNTSLYDCI--NSLLQ 492
DB      231 KLDGAVQDLRV-----AREKSGIQYQVQARRAL-----CVPPNPILY 270
QY      493 YFQNNRTLLLTANQTMGQTSQVDWKDHFLYCANAPLTFKDGTAALSCADYGA-EVF 551
DB      271 AMQVKTINL-----SSISFPAY-----NHRHRLY 296
QY      552 PFLAIGY-----KGDYSEAEALIMTFSLNTPYRADDPL--AQAKMEBAFLBEMRAF 603
DB      297 LTGFFGYTLGSLGMLLRAKAMRLLYLK--TEDPEYDVSKQMLTHLLDOFTNI 353
QY      604 QBRMA-----GMFOVTFARSLDEINRTAEDPIFATSYIVFLYISL 649
DB      354 KNTLLKXIBVPGVGLGGQGEKVHFTSLRQLEFEFTSVTVIVFHLATILILFAVT 413
QY      650 ALGYSYSWRVWVDSKATIGLGVAVVLGAVMAAMGFSYIGRSVLVLOVVPFLVLSV 709
DB      414 SCFRFD-----CIRKMCVAAFVGSIFLAVVSGGLLHIGV-PEVITIVANSPLILGV 467
QY      710 GADNIFIVLEYQRLPRRPGEPREVHIGBALGVAPSMILCSLSALICEFLGALPMPAV 769
```



```

RESULT 12
US-11-337-244-149
; Sequence 149, Application US/11337244
; GENERAL INFORMATION:
; APPLICANT: William Matthews
; APPLICANT: Mark Moore
; APPLICANT: Russell Phillips
; APPLICANT: Michael V. Wiles
; APPLICANT: Thadd C. Reeder
; APPLICANT: Robert G. Wiscotzkey
; APPLICANT: Keith D. Allen
; APPLICANT: Helen Barbault
; APPLICANT: Thomas J. Brennan
; APPLICANT: Catherine Guenther
; APPLICANT: Robert Klein
; APPLICANT: Christopher J. Kirk
; APPLICANT: Simon X. Xie
; APPLICANT: Qin Zhang
; APPLICANT: Agnes Chopin
; TITLE OF INVENTION: Targeted Gene Disruptions, Compositions
; FILE REFERENCE: R-DB-13
; CURRENT FILING DATE: 2006-01-20
; PRIOR APPLICATION NUMBER: US/11/337,244
; PRIOR FILING DATE: 2003-09-23
; PRIOR APPLICATION NUMBER: US 60/413,647
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,666
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,653
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,646
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,625
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,639
; NUMBER OF SEQ ID NOS: 237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 1182
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-337-244-149

Query Match      8.4%; Score 583.5; DB 6; Length 1182;
Beet Local Similarity 24.2%; Pred. No. 3,6e-38;
Matches 282; Conservative 154; Mismatches 433; Indels 297; Gaps 44;

QY 330 SFSSTHTLLGQ-----FPOG---MGTWVASWPLTIVLVIPIVVALAAGLVFTTELT 376
DB 20 SAPHILHSLAPLMLRAYFOGLFSLGCRIOKKGKVLFGIVAFGLALGLRAYAIE 79
QY 377 TDPEVMSAPNSQASSEKAFHDQHG-PFRFTNQVILTPNRSRYDSLLGPKNFSGI 435
DB 80 TDLBQLMWEVSGRSVSELHYTKKELGEBAAYTSQMLI---OTAHQEGGNVLTPE---A 131
QY 436 LDLDLLELLELOERLHLQVWSPEAQRIISLODIY---APLNPN-----TSLYDCC 486
DB 132 LD-----LHLOAALTASKVOVSUYKSWLANKICKSGVPLIEGMIERMIETKLPV 184
QY 487 INSLQYFQNNRTLLLTANQTLMGTSQVMDK---HFLYCANAPLT-----FKDGTALA 539
DB 189 ILTPIDCFMEGAK---LGGSAVILPERPIQWNTNIDPQQLBELGFPASLEGRRLDCA 241
QY 540 LSCMADYGAPV-----PPLA---IGYKGDYS----- 565
DB 242 QVGAQVYGRPCDDPDDPHCPSPAPNHSROAPVVAQELSGGCHGSGHKFMHNOBELLLG 301
QY 566 -----BAVALIMTF-----SLNNYPAGDPRPLAQAUKLMBEAFLEEMRFAFORBMG 609
DB 302 TARDLOGQLLRBAALOSTFLMSPROLYEHFRGDYGTIDIGNSEBOASVNLQWOKRFVQ 361

```

```

QY 610 MEQVTFTRERLE-----DEINRTTADBLPIFATSIYV---IFLYISLALSGYSWS 658
DB 362 LAQBALPANAQQIHAFSSTLIDILRAFSE-----VSTIRVVGGLMLAAYACVTMLRMD 417
QY 659 RVMVDSKATLGLGGAVALVGAVMAMGFSSYLGIRSSVLVQLQVPELVSGADNIFEV 718
DB 418 --CAOSQAVGLAGVLAVALVAGSLGICALGITFNATTQVLPFLGLGIGVDDIFLLA 475
QY 719 LEYORLPRRPBPBPBVHIGRALGRVAPSMGLCSSEALCFPLGALTMPAVRTALTSGL 778
DB 476 HAFKAP--PDTPLPERGEGCLRSCTGTSVALTSVNNVAFPMALVPIPALRAFSLQAI 533
QY 779 AVIIDLFLQSAFVALLSLSKROEASRLDVCCC-----VKQGE----- 817
DB 534 VGCNFAVMLVFPALISLDRRRRRLDVLCCFSSPCSAQVIMLPQELGDAVAPVGI 593
QY 818 -----LPP-----PG-----QGE-- 826
DB 594 AHLRTVQAFTHCEASSQHVVTILPQAHLLSPASDPLGSELSPGSGTRDLSQEBGTG 653
QY 827 -----LILGFFOKAYAPFLHMTTRGVLLPLALFGVSLYSNCHISVGLDQ 873
DB 654 PQACRPILCAHMTLAHFARVQFAPLLQTRAKALVLLFFGALLGSLYGATLVQDGLAL 713
QY 874 ELALPKDSYLLDYFLFLNRYFEVGAIFYVTTLGYNES--SEAGMNAICSSAGCUNNFSTQ 932
DB 714 TDVVPGRGTEHAFLSAQLRYPSL-YEVALVVOGGEFDYAHSGRAL-----FDLHQ 761
QY 933 KIQVATERPEOSYALIPAS-----SWVDDF-----IDMLTPSSCCRLYISGPNK 976
DB 762 RFSSL-----KAVLPPTAQAPRTWLAHYRSMLOGIOAFPODMASGITHSTRNGSED 816
QY 977 D----KPCPSIVNSLNCXKMSITM-----GSVPSVEQPFHKXLPWFLNDRPIKCPK 1026
DB 817 GALAVKLLIQTGNQGEPLDFSQLTRKLVDEKGLIP--ELFYMGLTWVSGDPL----- 869
QY 1027 GGLA-----YST---SVNLTSDQVLASRFMAVHKLXKSQDYTELR 1067
DB 870 -GLAASQANFYPPPEPWMLDKYDTTGENTLRIPAQOPEFFAFPFLLHGLQTAQFVEAIE 928
QY 1068 AABELAANI--TADLKVPGTDPAPEVFPYTTITNVFYEQLTILPBGFLMLSLC--LYPTF 1124
DB 929 GRRAACTEAGQGVNAVYSGSP-----LFMBQYLG--RRCFLVACILVCTF 976
QY 1125 AVSCLLGLDLRSGHLNLSIVMILVDTVGFMALMDISYNAVSLNLSAVGMSVEFVSH 1184
DB 977 LVCHALLLSLSPMTAGLI-VLVLAHMTVELFGIMFGIKLSALPVVILVASIGIGVEFVH 1035
QY 1185 ITRSFALSTKPTWLERAKEATISMGSAVFAGV---AMTNLPGLVLGLAKAQLIQIFFR 1241
DB 1036 VALGFLTSHSGRNLRPA-----SALEQTFAPVTDGAVSTLGLMLAGSNDFPIIRYFV 1090
QY 1242 LNLITLGLGLHGLVFLVLIISYGP 1267
DB 1091 VLVTLTGLGLHGLLPLVLIISLGP 1116

RESULT 13
US-60-772-265-1197
; Sequence 1197, Application US/60772265
; GENERAL INFORMATION:
; APPLICANT: BOURKHAROV, ANDREY A.
; APPLICANT: DU, ZIDIN
; APPLICANT: GUO, LIJANG
; APPLICANT: HRESKO, MICHELLE C
; APPLICANT: KOVALIC, DAVID K
; APPLICANT: ZHAOLONG, LI
; APPLICANT: LU, MAOLONG
; APPLICANT: MCCARTER, JAMES P
; APPLICANT: MILLER, NANCY M
; APPLICANT: VAUDIN, MARK
; APPLICANT: WILLIAMS, DERYCK J

```



```

1  APPLICANT: WU, WEI
2  TITLE OF INVENTION: IDENTIFICATION AND USE OF TARGET GENES
3  TITLE OF INVENTION: FOR CONTROL OF PLANT PARASITIC NEMATODES
4  FILE REFERENCE: NND1:002USp1
5  CURRENT APPLICATION NUMBER: US/60/772,265
6  CURRENT FILING DATE: 2006-02-10
7  NUMBER OF SEQ ID NOS: 1919
8  SOFTWARE: Patentin version 3.3
9  SEQ ID NO 1197
10 LENGTH: 891
11 TYPE: PRT
12 ORGANISM: Heterodera glycines
13 FEATURES:
14 OTHER INFORMATION: genomic DNA sequence=SeqID_981
15 US-60-772-265-1197

```

[illegible][illegible]

```

RESULT 14
US-60-772-265-241
/ Sequence 241, Application US/60772265
/ GENERAL INFORMATION:
/ APPLICANT: BOURKHAROV, ANDREY A.
/ APPLICANT: DU, ZIJUN
/ APPLICANT: GUO, LIANG
/ APPLICANT: HRESKO, MICHELLE C
/ APPLICANT: KOVALIC, DAVID K
/ APPLICANT: ZHAOLONG, LI
/ APPLICANT: LU, MAOLONG
/ APPLICANT: MCCARTER, JAMES P
/ APPLICANT: MILLER, NANCY M
/ APPLICANT: VAUDIN, MARK
/ APPLICANT: WILLIAMS, DERECK J
/ APPLICANT: WU, WEI
/ TITLE OF INVENTION: IDENTIFICATION AND USE OF TARGET GENES
/ TITLE OF INVENTION: FOR CONTROL OF PLANT PARASITIC NEMATODES
/ FILE REFERENCE: MN01:002USP1
/ CURRENT APPLICATION NUMBER: US/60/772,265
/ CURRENT FILING DATE: 2006-02-10
/ NUMBER OF SEQ ID NOS: 1919
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 241
/ LENGTH: 1356
/ TYPE: PRT
/ ORGANISM: Caenorhabditis elegans
/ FEATURE:
/ OTHER INFORMATION: C elegans public gene name=Y110A2A.L.8
/ FEATURE:
/ OTHER INFORMATION: genomic DNA homolog sequence=Segid_396
US-60-772-265-241

```

[illegible]

Oy 801 OQEARLVCCEVK-----POLP-----PRGQEGLL 828  
 Db 783 KQKAKRIDLAAYSRCNPNOMATISQSVPSNVSNSSRAELAGYEQKODEYRRHEPWTVG-- 840  
 Oy 829 LGFQKQAPAFELMHITRGVVLLEFLALPGVSLYSKCHISVGDQDELALPKDSYLLDPEL 888  
 Db 841 -GFLNKIITIPALKNNVYKACVLIIGTTNAVPLGYGVYTSITLGGELADVLPEHTPPAPFLR 899  
 Oy 889 FLNRYFEVGAPVYFY-----TLGYNPSSEAGNNAICSSAGCNNSFTQKIQYA 937  
 Db 900 AREQYFSF-YPMFAVLRGDKLDIPNQQLIEYRAQLG-----SSKFMILK 943  
 Oy 938 TEFPQSIVAIAPASSWVDFFIDML-----TPSS- 965  
 Db 944 AEGLQPY-----WMSMLRYVLOSJLDMALEKDLAAGFDLTNGNPILVNGEKPSBSM 996  
 Oy 966 -CCRLYIS-GPNKDKFCPESTVNSLCLKKNCMSITMGVSVPSEYQFHKYLP-WFLNDRPNI 1022  
 Db 997 IAARLVCSFGNTYN--CDORLQKMKVNEV-----INP--EGFTNYLTGFMND--NM 1044  
 Oy 1023 KCPKGGLAAYSTSVNLTSDDQVLASRFMAVYHKPLKNSQ-----DYTEALARARELA 1073  
 Db 1045 MYVYQSASFYPFRPGMEVNEKL--AKVPAABEPLAYSQMPFYQNDLIDTPALVKMIEIR 1102  
 Oy 1074 ANITADLKVPCTDPAFVFPPTTINVFEOYITLIPGLFMSLCLVTPFPAVSCLLGL 1133  
 Db 1103 ATCE-----EYSEGLSNHPSGIAFTFEQOYL-LRMNLFO-ALCIIL-ALAVFCVISTL 1153  
 Oy 1134 DLRSGLNLISIVMLL--VDTVGFMAIMDISYNAVSJNLVSAVGSVEFSHITSPAI 1191  
 Db 1154 MENPMATILINKIVITITIELGFMGLMKIKMNPISAVTLICAVIGVFETAHVELAF-L 1212  
 Oy 1192 STKPTMLERAKEATISMSGSAVEGAVAMTNLPGILVGLAKAQLIQIFFRNLMLITLGL 1251  
 Db 1213 TALGTIDQRLBESCLQHMFPVYVGHAISTFL-GVMVLVFSEPPVYVYFYTYMTLVALGV 1271  
 Oy 1252 LHGLVFLPVILSYVGPD-----VNPALEQKAEER 1283  
 Db 1272 FNGLCVLPIVILTVGPKPELPTDSSVLP.PPPPLNQYAEKS 1314  
 RESULT 15  
 US-11-360-355-133805  
 ; Sequence 133805, Application US/11360355  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boukharov, Andrey  
 ; APPLICANT: Gu, Zijiang  
 ; APPLICANT: Guo, Liang  
 ; APPLICANT: Kovallik, David  
 ; APPLICANT: Lu, Maolong  
 ; APPLICANT: McCarter, James  
 ; APPLICANT: Miller, Nancy  
 ; APPLICANT: Williams, Deryck  
 ; APPLICANT: Vaudin, Mark  
 ; APPLICANT: Wu, Wei  
 ; TITLE OF INVENTION: METHODS FOR GENETIC CONTROL OF HETERODERA INFESTATIONS  
 ; TITLE OF INVENTION: IN PLANTS AND COMPOSITIONS THEREOF  
 ; FILE REFERENCE: 38-21(53885)  
 ; CURRENT APPLICATION NUMBER: US/11/360,355  
 ; CURRENT FILING DATE: 2006-02-24  
 ; NUMBER OF SEQ ID NOS: 171306  
 ; SEQ ID NO 133805  
 ; LENGTH: 831  
 ; TYPE: PRT  
 ; ORGANISM: Heterodera glycines  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (794)..(794)  
 ; OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
 ; FEATURE:  
 ; OTHER INFORMATION: Coding regions on vCDNA, vCDNA=SeqID\_60228; Strand=-; Position=1  
 ; OTHER INFORMATION: -173..114-659,516-626,687-921,1040-1216,1282-1406,1470-1579,1627  
 ; OTHER INFORMATION: -1785..1833-1933,1984-2145,2192-2241,2295-2457,2503-2695,2750

```

OTHER INFORMATION: -2985,3164-3208
FEATURE:
OTHER INFORMATION: Homolog annotation: Hlt ID=NP_496761.1; Match_level="QueryCover=
99%, HitCoverage=86%, E_value=0.0, Identity=88%"; Hlt desc:pt:
OTHER INFORMATION: =atched Related (90.8 kD) [pfr-18] [Cenohadictis elegans]
OTHER INFORMATION: pfr||Y76683 hypothetical protein Y38f1a.3 - Cae
DS-11-360-355-133805

```

Query Match	6.3%	Score 437	DB 7	Length 831
Best Local Similarity	22.5%	Pred. Nc 1.6e-26		
Matches 189	Conservative 139	Mismatches 303	Indels 210	Gaps 30

```

OY 590 LUWEEAFLEENBARAQRBRAMGFOVTFERSTEDSINTTMEEDLIPATSYVIFPLYISTL 649
Db 21 KKEHSHSVYN--FAMNSRQDPLLKLAHTSSEGLVSEVRRTGIEVLPMLPMSIVILLFTYI 78
OY 650 A-----LGSY-----SSWSRVWVDSKATJGLGVAVVLGAVMAAMGFSSY 689
Db 79 TSLRKDOJRSKPEWALVGVFCPTVITSLKRDQJRSKPEWALVGVFCPIISVASSFGTLPW 138
OY 690 LGIR-----SLVILQVPPFLVTSVGDNIFFIPLERQRLPRRGEEREVHIGALGRV 743
Db 139 LGFBEPLGFEFLELVLVPPFLIAGVDVFIHCAHTD--PSKOLREVRADLLGSA 196
OY 744 APSMLGCSBAICFFLGALTPMPAVRTFATLSGLAVLIDFLQMSAFVALISLDSKOE 803
Db 197 GPSVTIISLTMWLSFTGIATPTPAIRFCLEFISAVYAYVLQYFFPTAAVWVIAQOREA 256
OY 804 ASRLDCCCVK--POELPPGQ-----GEGLLIGFFQKAYADPFLHW 843
Db 257 DERNAVYLCIRGVKPSRPKAVATLKKETNLMRLGSTNLMRLGSDLV-----PRYDVFWMSW 311
OY 844 ITR--GVULLFLPAL--FGVSLVSMCHISVGLDQELMPPDSYLDVDFELNR--YFEP 896
Db 312 TARIILAI SLTIYAFSAIGVAQIKVAQIKVGLTSEKIFLDDSPLEBYRLONNVIIFKEG 371
OY 897 GAVYFVTTLGYNFSSEAGMNAICSSAG--CNNFSFTOKIYATFEPPEOSYLAIPASS-- 952
Db 372 GOMAVFVNSPG--DLRKPMAVFNVNSPGDLRKKPSAIPRIMRILBERFQDSNGSVGSSTQM 428
OY 953 WVDPEIDMLTSSCCRLYISGPNKDKPCPSIVNSLNCJKNCSITMGSVR-----P 1003
Db 429 WLNTLYLPRTG-----LONNGSINRGSIDFRYKTYLDFEF 461
OY 1004 SVEOFHKLTPMF-----LNDPNIKCPKGLAAYSTSVULTSDQVLA SRFMAVHKP 1055
Db 462 SIPEHRSHSVLSGPKEDCLNERP-----SCINKFPDCLNERPSCINKFP--FSTG 511
OY 1056 LKNS---ODYEAALPAARELA---NITADLRKVPCTDAFVFFPTITTVFEOYIT 1107
Db 512 FQNAVMSDRVLVTLQRMQLADYSQMLNT-----YEDFMSVADOMYADQLIS 560
OY 1108 I-----LPEGLFMLSCLVPTFAVSCLLGL-----DLRSGLLNLSIWMILVDVFGFML 1158
Db 561 IPSYTIQVAPALCLCMTFVAPALCLCMTFVLMFPISITILPBTACVLSINIGVGLF 620
OY 1159 WDISYNAVSLINLVSAGMSVEFVSHI-----TSPFAISTKPYML----- 1198
Db 621 WSIDLDPISTMTTILMAIGLSVFAHISPHYKGMWTTLMAIGLSVDFVAHISPHYKGE 680
OY 1199 -----ERAKEITISMGSAVFAGVAMTNLPGLVLVGLAKAOLIQF----- 1238
Db 681 IEDSRERIRHRLSSITAMPLO--AALSTVLSIMLIVLIRAHYMOVAVKVVLLVVLVGLVHG 739
OY 1239 ---EFFRLNLTLLGLGLVFLPVILISYV-----GPDVN-----ALALEOKRA 1280
Db 740 LVEVVKVLLVVLVGLVGLVLLPVVYAAIFPQKSGATAKAPVPPLNRRRTAVTQKRRR 799
OY 1281 E 1281
Db 800 E 800

```

Search completed: March 22, 2006, 23:07:41  
Job time : 63 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 23:07:52 ; Search time 25 Seconds  
(without alignments)  
1525.028 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909  
Sequence: 1 MABAGRGWMLWALLRLAQ.....GSIKGAGAIENFLPNNGROF 1332

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 28622889 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications\_AA\_New:\*  
1: /cgn2\_6/prodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/prodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
3: /cgn2\_6/prodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
4: /cgn2\_6/prodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/prodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
6: /cgn2\_6/prodata/1/pubppaa/US11\_NEW\_PUB.pep:\*  
7: /cgn2\_6/prodata/1/pubppaa/US11\_NEW\_PUB.pep:\*  
8: /cgn2\_6/prodata/1/pubppaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6896	99.8	1332	US-11-242-459-9	Sequence 9, Appl1
2	2402.5	34.8	1278	US-10-995-561-952	Sequence 952, App
3	942	13.6	1394	US-11-242-459-8	Sequence 8, Appl1
4	195.5	2.8	1087	US-11-242-459-17	Sequence 17, Appl1
5	195.5	2.8	1392	US-11-242-459-40	Sequence 40, Appl1
6	191	2.8	127	US-11-096-568A-6778	Sequence 6778, Ap
7	173	2.5	85	US-11-096-568A-6779	Sequence 6779, Ap
8	143	2.1	445	US-11-242-459-35	Sequence 35, Appl
9	143	2.1	456	US-11-242-459-34	Sequence 34, Appl
10	141.5	2.0	1048	US-10-392-234A-16	Sequence 16, Appl
11	140.5	2.0	1046	US-10-392-234A-16	Sequence 16, Appl
12	134	1.9	1067	US-10-467-657-2648	Sequence 2648, Ap
13	127	1.8	1048	US-10-392-234A-14	Sequence 14, Appl
14	126.5	1.8	200	US-11-242-459-16	Sequence 16, Appl
15	126	1.8	801	US-10-793-626-2020	Sequence 2020, Ap
16	121.5	1.8	1043	US-10-392-234A-34	Sequence 34, Appl
17	120.5	1.7	2426	US-11-203-806A-11	Sequence 11, Appl
18	119	1.7	1048	US-10-392-234A-20	Sequence 20, Appl
19	117.5	1.7	1049	US-10-392-234A-12	Sequence 12, Appl
20	115.5	1.7	914	US-11-072-512-2923	Sequence 2923, Ap
21	113	1.6	1034	US-10-392-234A-30	Sequence 30, Appl
22	112.5	1.6	697	US-11-082-389-362	Sequence 362, App
23	109.5	1.6	758	US-11-096-568A-30412	Sequence 30412, A
24	109.5	1.6	772	US-11-096-568A-30411	Sequence 30411, A
25	109.5	1.6	827	US-11-096-568A-30410	Sequence 30410, A

26	109.5	1.6	1049	7	US-11-098-686-11239	Sequence 11239, A
27	109	1.6	1032	6	US-10-392-234A-67	Sequence 67, Appl
28	108.5	1.6	648	7	US-11-087-099-4135	Sequence 4135, A
29	106	1.5	443	7	US-11-087-099-11457	Sequence 11457, A
30	105.5	1.5	783	7	US-11-082-389-354	Sequence 354, App
31	105	1.5	324	7	US-11-087-099-11792	Sequence 11792, A
32	105	1.5	1080	7	US-11-096-568A-21723	Sequence 21723, A
33	105	1.5	1097	7	US-11-096-568A-21722	Sequence 21722, A
34	105	1.5	1128	7	US-11-096-568A-21721	Sequence 21721, A
35	104.5	1.5	315	7	US-11-087-099-9820	Sequence 9820, Ap
36	104.5	1.5	757	7	US-11-087-099-4362	Sequence 4362, Ap
37	104	1.5	771	7	US-11-096-568A-21509	Sequence 21509, A
38	104	1.5	826	7	US-11-096-568A-21508	Sequence 21508, A
39	104	1.5	826	7	US-11-096-568A-21507	Sequence 21507, A
40	103.5	1.5	251	7	US-11-087-099-1098	Sequence 1098, Ap
41	103.5	1.5	331	7	US-11-087-099-8831	Sequence 8831, Ap
42	103.5	1.5	2312	7	US-11-126-313-34	Sequence 34, Appl
43	103	1.5	463	7	US-11-087-099-3980	Sequence 3980, Ap
44	102.5	1.5	475	7	US-11-087-099-2173	Sequence 2173, Ap
45	102	1.5	324	7	US-11-087-099-5717	Sequence 5717, Ap

## ALIGNMENTS

RESULT 1									
US-11-242-459-9									
Sequence 9, Application US/11242459									
Publication No. US20060035835A1									
GENERAL INFORMATION:									
APPLICANT: TANIYAMA, Yoshio									
APPLICANT: KITTA, Shunbun									
TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof									
FILE REFERENCE: 2701USDP									
CURRENT APPLICATION NUMBER: US/11/242,459									
CURRENT FILING DATE: 2005-10-03									
PRIOR APPLICATION NUMBER: US/10/239,316									
PRIOR FILING DATE: 2002-09-19									
PRIOR APPLICATION NUMBER: PCT/JP01/02279									
PRIOR FILING DATE: 2001-03-22									
PRIOR APPLICATION NUMBER: JP2000-088595									
PRIOR FILING DATE: 2000-03-24									
NUMBER OF SEQ ID NOS: 59									
SEQ ID NO 9									
LENGTH: 1332									
TYPE: PRT									
ORGANISM: Human									
US-11-242-459-9									
Query Match									
Best Local Similarity 99.8%; Score 6896; DB 7; Length 1332;									
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;									
QY	1	MABAGRGWMLWALLRLAQSEPTTIHQGYCAFYDEGCKNPELSGLMTLSNVCLSN	60						
DB	1	MABAGRGWMLWALLRLAQSEPTTIHQGYCAFYDEGCKNPELSGLMTLSNVCLSN	60						
QY	61	TPARKITGDHLIILOKICPRPLTYGNTQACCAKQVLSBASISITKALLTRCPACSDNF	120						
DB	61	TPARKITGDHLIILOKICPRPLTYGNTQACCAKQVLSBASISITKALLTRCPACSDNF	120						
QY	121	VNIHCNTPSPNOSTLFTNTRVQGLGAGOLPAVVAEAFQHSFAROSYDCSRVVPAA	180						
DB	121	VNIHCNTPSPNOSTLFTNTRVQGLGAGOLPAVVAEAFQHSFAROSYDCSRVVPAA	180						
QY	181	ATLAVGTMGVYSALCNARWLNFGODTGNGLAPDITFHLPEQAVSGGIQPINEG	240						
DB	181	ATLAVGTMGVYSALCNARWLNFGODTGNGLAPDITFHLPEQAVSGGIQPINEG	240						
QY	241	ARCNEGQDDVATVCSQDCDCAASCPAIPARQALDSTYTLGQMPSTVLIIILCSVPAVVTI	300						
DB	241	ARCNEGQDDVATVCSQDCDCAASCPAIPARQALDSTYTLGQMPSTVLIIILCSVPAVVTI	300						

QY 301 LTVGRVADPARDKSKVNDPKKSTLSLSDKLSFSTHTLLGQFPQGMGTWASWPLTTLVLSV 360  
 DB 301 LTVGRVADPARDKSKVNDPKKSTLSLSDKLSFSTHTLLGQFPQGMGTWASWPLTTLVLSV 360  
 QY 361 TVVVALAAGVTELTTPDVELMSAPNSQARSSEKAPHDHGFFPTNQTIVITAPRSRY 420  
 DB 361 TVVVALAAGVTELTTPDVELMSAPNSQARSSEKAPHDHGFFPTNQTIVITAPRSRY 420  
 QY 421 RVDSLLGPKNFGSLDLDLLELLELOERLRLQVMSPEAGRNISLQDICYAPLNDPT 480  
 DB 421 RVDSLLGPKNFGSLDLDLLELLELOERLRLQVMSPEAGRNISLQDICYAPLNDPT 480  
 QY 481 SLVDCINSLLQYFQNNRTLLLTITANQTMGQTSQVMDKHFLYCANAPLTFKQGTALAL 540  
 DB 481 SLVDCINSLLQYFQNNRTLLLTITANQTMGQTSQVMDKHFLYCANAPLTFKQGTALAL 540  
 QY 541 SCMDVGAVPFPLAIGYKKGKDYSEAEALIMFSLNNYPAQDPRLAQAKWEAEALDEM 600  
 DB 541 SCMDVGAVPFPLAIGYKKGKDYSEAEALIMFSLNNYPAQDPRLAQAKWEAEALDEM 600  
 QY 601 RAFORRMAGFOVTTFAERSLDEINRTTAEPLIPATSYIVFLYISLALGSSMSRY 660  
 DB 601 RAFORRMAGFOVTTFAERSLDEINRTTAEPLIPATSYIVFLYISLALGSSMSRY 660  
 QY 661 MVDSKATLIGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIVLE 720  
 DB 661 MVDSKATLIGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIVLE 720  
 QY 721 YQRLPRRPREPREVHIGRALGRVABSMILCSLSEALCFPLGALTMPNARTPALTGLAV 780  
 DB 721 YQRLPRRPREPREVHIGRALGRVABSMILCSLSEALCFPLGALTMPNARTPALTGLAV 780  
 QY 781 IIDLFLQMSAFVALLSLDSKROBASRLDVCCCVKPOELPPQGGEGELIGFQKAYAPPL 840  
 DB 781 IIDLFLQMSAFVALLSLDSKROBASRLDVCCCVKPOELPPQGGEGELIGFQKAYAPPL 840  
 QY 841 LHMITRGVLLFLALFGVSLYSWCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900  
 DB 841 LHMITRGVLLFLALFGVSLYSWCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900  
 QY 901 YFPTTLGYNSSBAGNNALCSSAGCNNSFTQKIQTATEPREGSYALATPASSVVDFTDM 960  
 DB 901 YFPTTLGYNSSBAGNNALCSSAGCNNSFTQKIQTATEPREGSYALATPASSVVDFTDM 960  
 QY 961 LTPSSCCRLYISGPNKDFCPTVNSLNCIKNCMSITMGSVRSVEQFHKYLPMFLNDRP 1020  
 DB 961 LTPSSCCRLYISGPNKDFCPTVNSLNCIKNCMSITMGSVRSVEQFHKYLPMFLNDRP 1020  
 QY 1021 NIKCPKGLAAYSTSVNLTSQGVLASRPMAYHKPKNSQDYTEALRAARELANITADL 1080  
 DB 1021 NIKCPKGLAAYSTSVNLTSQGVLASRPMAYHKPKNSQDYTEALRAARELANITADL 1080  
 QY 1081 RXPVGTDPAPFEVFPYITTVFYEQYLTILPEGLFMTSLCLVPFPFASCLLGLDLSGGL 1140  
 DB 1081 RXPVGTDPAPFEVFPYITTVFYEQYLTILPEGLFMTSLCLVPFPFASCLLGLDLSGGL 1140  
 QY 1141 NLLSTMIILVDTGFMALMDISYNAVSLINLVASGVMSYFVSHIRSFALSTKPTWLB 1200  
 DB 1141 NLLSTMIILVDTGFMALMDISYNAVSLINLVASGVMSYFVSHIRSFALSTKPTWLB 1200  
 QY 1201 AKEATISMGSAVPAVAMTNLPGILVLGLAKAQLIOPFERLNLITLLGLHLGVLFLV 1260  
 DB 1201 AKEATISMGSAVPAVAMTNLPGILVLGLAKAQLIOPFERLNLITLLGLHLGVLFLV 1260  
 QY 1261 ILSYVDPVNPALALQKAAEEVAAVMAVASCNHPSRVSTADNITYNHSFBSSIGAGA 1320  
 DB 1261 ILSYVDPVNPALALQKAAEEVAAVMAVASCNHPSRVSTADNITYNHSFBSSIGAGA 1320  
 QY 1321 ISNFLPNNGRQF 1332  
 DB 1321 ISNFLPNNGRQF 1332

RESULT 2  
 US-10-995-561-952  
 ; Sequence 952, Application US/10995561  
 ; Publication No. US20050272054A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
 ; FILE REFERENCE: CLO01559  
 ; CURRENT APPLICATION NUMBER: US/10/995,561  
 ; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FaastSeq for windows Version 4.0  
 ; SEQ ID NO 952  
 ; LENGTH: 1278  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 US-10-995-561-952  
 Query Match 34.8%; Score 2402.5; DB 6; Length 1278;  
 Best Local Similarity 39.8%; Pred. No. 2,6e-181;  
 Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;  
 QY 7 RGMILLMALLRLAQSEPYTTHIQPGYCAFYDECG-----KNPELSGLMTLSNVSCLS 59  
 DB 4 RLALGLLTLTLCPAQVFSQ-----SCWYGCGLIAYGDKRNCVEYSG----- 46  
 QY 60 NTPAKTIGDHIILLOKICPRLYTGPNTQACCSAQVLSLEASITLALLTRCPACSDN 119  
 DB 47 --PPLPDGDYDLVQELCPGFFFG-NVSLCCDVQDLTKDNLOLPQFJSRCPSCFYN 103  
 QY 120 FVNLCHNTGSPNQLFLINVTRE----VAQLGAGQLPAVAAEAFYQHSFAESYDSCSRV 175  
 DB 104 LNLCELTCSRQGFPLNVTATEDYVDVPTNQTNTVKEIQYVYGQGFANMVAACDV 163  
 QY 176 RVPAATLAAGTGCVGSALCNAQRWLNFGQDGTNGIAPLDIT-----PHLLPQQA 228  
 DB 164 EAPSSNDALGILCGCKDA--CNATNMIEMFNKONGAPFTITVFSDFPVH----- 215  
 QY 225 VSSGIQPLNEGVARBNESQGDVATCSQDCAAGCPALARP-----ALDSTFYLG 279  
 DB 216 --GMEPMNATKGCDSDEVETAPSCQDCSIVGPKPQPPPPAPMTILGLDMYIM 272  
 QY 280 QMPGSLVLIILCSYFAVVTILL-----VGRVAPARDKSKMVDPKKSTLS 326  
 DB 273 WITWMAFLVFGAFATWCKRKYFVSEYTPIDSNIAFSV--NADKGE-----ASCC 324  
 QY 327 DKLSTHTLLGQFPQGMGTWASWPLTTLVLSVIVVALLAGVTELTTPDVELMSAP 386  
 DB 325 DVSAAAFEGCLRLFTBRGSPFCVRNPGCVIFPSLVIITACSGGLFVFTVTPVLMASAP 384  
 QY 387 NQOASSEKAFHQHCGPFRFTNQVILTPAPNSSYRYSLLGPK--NFGSLIDLDLLELL 445  
 DB 385 SQQAILEKEYPQHGHPFRFTQILIRAPLTDKHIYQPSGADVFPFAPDIQILHVL 444  
 QY 446 ELQERLRLQVMSPEAGRNISLQDICYAPLNDPTSLVDCINSLLQYFQNNRTLLT 505  
 DB 445 DLQIAIEN--ITASDNETVITQDCLAPLSYNT--NCTILSTANTFQNSHSLDHKK 499  
 QY 506 NOTLMGQTSQVMDKHFLYCANAPLTFKQGTALALSCMDYGAVPFPLAIGYKGDYS 565  
 DB 500 GDDFF--VYADYHHPFLYCVAPASLNDTSLHDPCLTGFGAPVPMVLGQYDDQYN 556  
 QY 566 EAEALIMFSLNNYPAQDPRLAQAKWEAEALFEKRAQORRMAGFOVTTFAERSLEDEI 625  
 DB 557 NTAIVITFPVNNYNDTEKQRAQAWKEKFNFPVKNK--PMLTSTFAERSIEDEL 613  
 QY 626 NRTTAEPLIPATSYIVFLYISLALGSSMSRYVNDPKKATLIGGVAVVLGAVMAAMG 685  
 DB 614 NRESDSVFTVTVISTAIHFYISLALGHIKSCRLLVDSKSLGSLAGILIVLSVACSLG 673



QY 686 FFSYGISSVLIVLVFPELVSVADNFI FLEKORLPRRPGREAEHIGALORVP 745  
Dh 674 VFSYIGLEPLTIVIEVIPVLAVGVDNIFILVQAYQRBRLQGETLIDQOLGRVIGEVAP 733  
QY 746 SMLCSLSEBACIFPGALTPMPBAYRTFALTSGLAVIDLELOMSAFVALLSDSKROBAS 805  
Dh 734 SMFLSSFSSETVAFPIGALSVMPDAVHTFSLFAGLAVPIDLTQTCVSLIGLIDIKROEKN 793  
QY 806 RLDDVCCYKPOGLPPPGQ-GEGLLIGFPQKAYAPPLMHHTTRGVULLIFLALFGVSLYSM 864  
Dh 794 RLIDIFCCVRGADGTSVQASEBCLPRFFFNKSYSLPLKXOMMBPIVIAIFGVLSFISIAVL 853  
QY 865 CHISVGLQEOBALPXDYSLLDYELFLNRYEEVGAAPYFTTLTGYNSSSEAGMAICSSAG 924  
Dh 854 NKVDIGLQOSLMPDSTWVDIFKXISQTLHGPPRYFVLEESHDITSSKQOMVCGCMG 913  
QY 925 CNNSFPTOKIQYATFPEOSYLDAIPASSWVDFIDWLP--SSCCRLYISGPNKDKCPST 983  
Dh 914 CNDISLVQOIFPAQOLDNWTIRIGFAPSSWIDDYFPMVKRQSSCCR---DNITDQCNMS 970  
QY 984 VNSLNCLEKNCMSIT-MGSVRSVEQFHKYLPMFLANDRPNIKCPKGLAAYSTSVN--LTS 1040  
Dh 971 VVDPACV-CRPLTEPGKRCPOGCDFMRLPMFLSNPNPKCGKGHAAYSSAVNTILGH 1029  
QY 1041 DGOVLASPFMYHKLEKNSODYTEALRAREBLAANTALARKVCGTDPAPEVPPYITMY 1100  
Dh 1030 GTRGATIFEMTHIVLQTSADRIDALKKARLLASNT-ETMGNS--NRYFVSVFVY 1086  
QY 1101 FVEOYLTLIPBGLFMLSCLVPTFPVASCILLGIDRSGLNLTLSYMIIVDTGVFALMD 1160  
Dh 1087 FYEQULTIIDPTIFMLGVSIGAFLVTWVLLGCELMASAIMCATIMVLVNMGGVWIMNG 1146  
QY 1161 ISYNAVSLINLVASVGMSEYFVSHITRSFAISTKPTWLEBAKEATTSMGSAVFAVAMTN 1220  
Dh 1147 ISLNAVSLVNLVMSGSISEVFCSHITRAFTVSMKGSRVRAEBALAHMSSSVFSGITLTK 1206  
QY 1221 LRGILVGLAKAQLQIIFPERLNLITLIGLHGVLFLVIIISVYGPDPNPA 1272  
Dh 1207 FGGIVLVLPASSQIFQIIFPRMYLMLVNLGATHGILFLFVLLSLYDPSVNKA 1258

```

: RESULT 3
: US-11-242-459-8
: Sequence 8, Application US/11242459
: Publication No. US20060035835A1
: GENERAL INFORMATION:
: APPLICANT: TANIYAMA, Yoshio
: APPLICANT: KITA, Shunbun
: APPLICANT: SATOMI, Tomoko Komiya
: TITLE OR INVENTION: Novel Protein, Process for Producing The Same And Use Thereof
: FILE REFERENCE: 2703USOP
: CURRENT APPLICATION NUMBER: US/11/242,459
: CURRENT FILING DATE: 2005-10-03
: PRIOR APPLICATION NUMBER: US/10/239,316
: PRIOR FILING DATE: 2002-09-19
: PRIOR APPLICATION NUMBER: PCT/JP01/92279
: PRIOR FILING DATE: 2001-03-22
: PRIOR APPLICATION NUMBER: JP2000-088595
: PRIOR FILING DATE: 2000-03-24
: NUMBER OF SEQ ID NOS: 59
: SEQ ID NO 8
: LENGTH: 194
: TYPE: PRT
: ORGANISM: Human
: US-11-242-459-8

```

```

Only Match      13.6%  Score 942; DB 7; Length 124;
Best Local Similarity 99.5%  Pred. No. 4, 8e-67;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0

Qy      615  FTAERSLDEINFTVADLPIFATSYIVIFLYISIALGSSYSWSRWMDSTATTGLGVA 674
      1  FMAERSLDEINFTVADLPIFATSYIVIFLYISIALGSSYSWSRWMDSTATTGLGVA 60

```

QY	675	VVLGAMAMGEFSYLGRSSVLIQVVPFLVLSVGDNIIFVLEVORLPRRGEEREV	734
Db	61	VVLGAMAMGPFSSYLGRSSVLIIQVVPFLVLSVGDNIIFVLEVORLPRRGEEREV	120
QY	735	HIGRALGRVAPSMILCSLSEACFPGLGALTMPAVRPTALTSGLAVIIDFLQMSAEVAL	794
Db	121	HIGRALGRVAPSMILCSLSEACFPGLGALTMPAVRPTALTSGLAVIIDFLQMSAEVAL	180
QY	795	LSLDSKROEASRLD	808
Db	181	LSLDSKROEASRLD	194

```

RESULT 4
US-11-242-459-17
: Sequence 17, Application US/11242459
: Publication No. US20060035835A1
: GENERAL INFORMATION:
: APPLICANT: TANIYAMA, Yoshio
: APPLICANT: KITAI, Shunbun
: APPLICANT: SATOMI, Tomoko KomiYama
: TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof
: FILE REFERENCE: 2703USOP
: CURRENT APPLICATION NUMBER: US/11/242,459
: CURRENT FILING DATE: 2005-10-03
: PRIOR APPLICATION NUMBER: US/10/239,316
: PRIOR FILING DATE: 2002-09-19
: PRIOR APPLICATION NUMBER: PCT/JP01/02279
: PRIOR FILING DATE: 2001-03-32
: PRIOR APPLICATION NUMBER: JP2000-088555
: PRIOR FILING DATE: 2000-03-24
: NUMBER OF SEQ ID NOS: 59
: SEQ ID NO 17
: LENGTH: 1087
: TYPE: PRT
: ORGANISM: Human
: US-11-242-459-17

```

Query Match	2.8%	Score 195.5	DB 7	Length 1087
Best Local Similarity	19.5%	Pred. No. 5,56-07		
Matches 200	Conservative 123	Mismatches 320	Indels 379	Gaps 477
QY	576	LNNYPADPRLAQAWEBAFLBEMRAFORRMAMFOVTFARSLD-----	623	
Db	100	LPNYSVDDR-----WEE-----GQAKFGSFVVTYVAMLAQSTSKVLYGSTD	144	
QY	624	-----EINRTAED--LPFATSYIVFLYISLALGISTSMSRWVDSKATIGLGG--VAVV	676	
Db	145	LFDEYERRRTNNMDMLAFISSSCDIALVYLITSCSVRLFSGGI-----ASIGLSCLVALF	199	
QY	677	IGANMAMGFGSYGIGRSSLVILQVBPVLVSCADNIFIFVLEYORLPR--RPGEPREHV	735	
Db	200	LYHVVFGI---QYIGLNG---VAAFVTVGSDVDVFINTYRQATHLEDQLKMIH	251	
QY	736	IGRALGRVAPSMILCLSEALICEFLGATITMPAVRTALYSGLAVI---LDFLLOMSAFV	792	
Db	252	TVQYAGK---ATPFSTLTAAVAANYFSQIPANHDGLFMSSLVSCOMLAVLTMRAAL	308	
QY	793	ALISLDSKROBASHLDVC--CYAKPOELPPPGQ-----	823	
Db	309	GLWLSLYLAPLESSQOTSCNONCSKRTSLHPDPVFAFAEOVGSGSPAGGPIPYLDDIPLL	368	
QY	824	-----GSELILGFQKAVAP-----	850	
Db	369	EVEEPEVSLBEGDVLVSVPBGLQFAPSNTSGRCHLIYLOELHHVWLSAIVSRWYIV	428	
QY	851	LLFLALFGVSL-----YSM-----CHISVGLDOE--	874	
Db	429	GLPFSIILISVFASTRLRAPASRAELRPDPNTIOVLDDKYNLSABEGISCTIGSLRQEK	488	
QY	875	-----LALPK-----DSYLLDY--FLFLNRYFEGAPVYFVTTLGNF	910	

Db 489 PHSLONNIRTSLEKRRGSGVPMASRPEATLODPCTVYISKYSGHRA--VYRLSLNA 546  
Qy 911 SSEAGMNAICSSAG--CNNF-----SFTOKIQVATEPPEOSYLAIPASSWDDFIDWL 961  
Db 547 SLPAWQAVSPGDGEVPSFOVYRAPFGNFTKLTACMSIVGLLQAASPSRKKM-----L 600  
Qy 962 TPSSC-----CRLYIS--GP 974  
Db 601 TTLACDARKGKWFPSFYVATKEOQHTKLYFAQSHKPPFHGRVCMAPPGCLSSSDGP 660  
Qy 975 NKDKFCPSTVNSLNCIKNCMSITMG-----SVRPSVEQPHKYLPMF-----LN- 1017  
Db 661 TKGFF---FVPSKVPKALSLATFGNPCVNTGCGKPAVRPLVDGAMFVFGIIGVNR 717  
Qy 1018 ---DRPNIKCPKGGIAAYSTSVNLTS-----GQVL 1045  
Db 718 TRQVNNHVGDP--GSVYDSSFDLFKEIGHLCHLCKAIAANSELVKPGAOCLPSGYSI 775  
Qy 1046 ASRFMAVYHKKPLKNSOD-----YTEALPAARELANITADLRKVGTDPAPFVF--P 1094  
Db 776 SSFLQMLHPECKELPEPNLLPGQLSHGAVVREGHVQWISMAFESTTYKGS--SFQYSD 834  
Qy 1095 YTTNVFYEQYLTILPEG-----LFM-----LSLCL--VP 1122  
Db 835 YLRMESFLQOQLALPEGSVLRGRFOTCEHMKQIFMEIVGVSALCGVLVSLICVAAYA 894  
Qy 1123 TPAVSCLLGLDRLSGNLNLSTVMILVDYGFPMALWDISYNAVSLINLVASVGSVEFV 1182  
Db 895 VFTTHLL---LPLVLSILIGIVLV--TIMYSGEM--GAVERAISILVGSVDYC 947  
Qy 1183 SHITTSFAIS-----TKPTWLEBAKEATISGSAVFGVAMTNLPGI----- 1224  
Db 948 VHLVEGYLLAGENLPHQAEADARTQROW--RTLEAVRHVGAIVSSALTTVIATVPLFC 1005  
Qy 1225 LVTLGAKAQLIQIFFFRNLITLLGLHGVFLPVILSYGPDVNPALALQKRAE--E 1282  
Db 1006 TIAPFAK-----FGKIYALNTGVSILYTLTVSTALGLIAPPS-----SFRTRTSFLK 1053  
Qy 1283 AVAAVAVVA 1290  
Db 1054 ALGAVILIA 1061  
RESULT 5  
US-11-242-459-40  
; Sequence 40, Application US/11242459  
; Publication No. US20060035835A1  
; GENERAL INFORMATION:  
; APPLICANT: TANIYAMA, Yoshio  
; APPLICANT: KITA, Shunbun  
; APPLICANT: SATOMI, Tomoko Komiyama  
; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use thereof  
; FILE REFERENCE: 2703USOP  
; CURRENT APPLICATION NUMBER: US/11/242,459  
; CURRENT FILING DATE: 2005-10-03  
; PRIOR APPLICATION NUMBER: US/10/239,316  
; PRIOR FILING DATE: 2002-09-19  
; PRIOR APPLICATION NUMBER: PCT/JP01/02279  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: JP2000-088595  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 59  
; SEQ ID NO 40  
; LENGTH: 1392  
; TYPE: PRT  
; ORGANISM: Human  
US-11-242-459-40  
Query Match 2.8%; Score 195.5; DB 7; Length 1392;  
Best Local Similarity 19.5%; Pred. No. 7.9e-07;  
Matches 200; Conservative 129; Mismatches 320; Indels 379; Gaps 47;  
Qy 576 LNNYPAGDRILAQAKIMEBAFLBEMRAFORRMAGMFQVTTABRSLED----- 623

Db 405 LPNYSVDDR-----WEB-----QRAKFGSFVTVYAMMLKKOSTSKYQVLYGCTD 449  
Qy 624 ---EINRTTAEED--LPFATSYIVIFLYISLALGSYSWSRWVDSKATIGLGG--VAVV 676  
Db 450 LFDYEVRRTPNNMDLAFISSICIALYIILTSCVSPISFGI-----ASIGSLCLVALF 504  
Qy 677 LGAVNANGFFSYLGIRSSVILQVPPFLVSVGADNIFIVLEFYQILPR--RPGEPREVN 735  
Db 505 LYHNVFEGI---QYGLILNG---VAAFIYIGIDVDVFEVINTYRQATHEEDPOLRWIH 556  
Qy 736 IGRALGRVAPSMILCSLEAICFPLGALTPMPAVRTALTSGLAVI---LDELQMSAFV 792  
Db 557 TVQTNAGK--ATFFTSLTAAIYANVPSQIPAVHDFGLFMSLIYSCCLAVLVTMPAL 613  
Qy 793 ALLSDSKROEASRLDVC--CCVKQDELPPGQ-----FLHWI-----TRGYVL 850  
Db 614 GLMSLYLAPLBESSCQTSCHONCSRTSLHPGDVFAPEQYGSBAQGPPIFYLDODIPLL 673  
Qy 824 -----GEGILLFPQKAYAP-----FLHWI-----TRGYVL 850  
Db 674 EYEEBPVBLELGDVLSVSPBGLQPAANTSGRHLIVQLQELHWMVLSAVKSRWIV 733  
Qy 851 LFLALFGVSL-----YSM-----CHISVGLDOE- 874  
Db 734 GLFVSLILSLVFSRLRPAARAPLFRPDNTNIQVLLDLKTNLSAEGISCTTCSGLPQK 793  
Qy 875 -----LALFK-----DSYLLDY--FLELNRYPEVGAPVYFVTTLAGYNF 910  
Db 794 PHSLONNIRTSLEKRRGSGVPMASRPEATLODPFGTYISIKYSGHRA--VYRLSLNA 851  
Qy 911 SSEAGMNAICSSAG--CNNF-----SFTOKIQVATEPPEOSYLAIPASSWDDFIDWL 961  
Db 852 SLPAWQAVSPGDGEVPSFOVYRAPFGNFTKLTACMSIVGLLQAASPSRKKM-----L 905  
Qy 962 TPSSC-----CRLYIS--GP 974  
Db 906 TTLACDARKGKWFPSFYVATKEOQHTKLYFAQSHKPPFHGRVCMAPPGCLSSSDGP 965  
Qy 975 NKDKFCPSTVNSLNCIKNCMSITMG-----SVRPSVEQPHKYLPMF-----LN- 1017  
Db 966 TKGFF---FVPSKVPKALSLATFGNPCVNTGCGKPAVRPLVDGAMFVFGIIGVNR 1022  
Qy 1018 ---DRPNIKCPKGGIAAYSTSVNLTS-----GQVL 1045  
Db 1023 TRQVNNHVGDP--GSVYDSSFDLFKEIGHLCHLCKAIAANSELVKPGAOCLPSGYSI 1080  
Qy 1046 ASRFMAVYHKKPLKNSOD-----YTEALPAARELANITADLRKVGTDPAPFVF--P 1094  
Db 1081 SSFLQMLHPECKELPEPNLLPGQLSHGAVVREGHVQWISMAFESTTYKGS--SFQYSD 1139  
Qy 1095 YTTNVFYEQYLTILPEG-----LFM-----LSLCL--VP 1122  
Db 1140 YLRMESFLQOQLALPEGSVLRGRFOTCEHMKQIFMEIVGVSALCGVLVSLICVAAYA 1199  
Qy 1123 TPAVSCLLGLDRLSGNLNLSTVMILVDYGFPMALWDISYNAVSLINLVASVGSVEFV 1182  
Db 1200 VFTTHLL---LPLVLSILIGIVLV--TIMYSGEM--GAVERAISILVGSVDYC 1252  
Qy 1183 SHITTSFAIS-----TKPTWLEBAKEATISGSAVFGVAMTNLPGI----- 1224  
Db 1253 VHLVEGYLLAGENLPHQAEADARTQROW--RTLEAVRHVGAIVSSALTTVIATVPLFC 1310  
Qy 1225 LVTLGAKAQLIQIFFFRNLITLLGLHGVFLPVILSYGPDVNPALALQKRAE--E 1282  
Db 1311 TIAPFAK-----FGKIYALNTGVSILYTLTVSTALGLIAPPS-----SFRTRTSFLK 1358  
Qy 1283 AVAAVAVVA 1290  
Db 1359 ALGAVILIA 1366  
RESULT 6



Matches	69, Conservative	41, Mismatches	106, Indels	56, Gaps	12, Gaps
Qy	576 LNTNPAGDPRLAQAKLWEAFLEEKRAFORRMAGNFOYFTTAERSLED-----	623			
Db	100 LPNYSYVDR-----WEE-----QRAKFOGFVTTYAMAKKOSTSKQVLYGGTD	144			
Qy	624 ----EINRTTAE-D-LPIFATSYIVIFPYISIALOSYSWSRWVDSKATLGLG-VAVV	676			
Db	145 LFDYEVRTTFNNMDMLAFISSCIAALVYILTSCVFLSFGI-----ASIGSLCLVALF	199			
Qy	677 LGAVMAAGFFSFSYIGIRSSLVILQVPPVLVSVGDNIFIFVLEYQRLPR-RPGEPREHV	735			
Db	200 LYHVVFGI---QYGLILNG-----VAAEVYIGIGVDVFEVFNTRYQATHTLEDPLRMIH	251			
Qy	736 IGRALGRVAPMLTCLISEAICFFGLGATLPMPAVTPALTSGLAVI---LDFILQMSAFV	792			
Db	252 TVQNRGK---ATFFSLTTLTAAYANVNSQIDPAVHDFGLFMSLIVSCMLAVLYTMPAAL	308			
Qy	793 ALLSIDSKRQEARSLDVC--CCVKKQELPPPG	822			
Db	309 GLWSLYLAPLLESSCQTSCHQNSRRTYSHPFG	340			

```

RESULT 10
US-10-392-234A-18
; Sequence 18, Application US/10392234A
; Publication No. US20050255538A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia and Upjohn Corporation
; APPLICANT: Buxser, Steven
; APPLICANT: Poole, Keith
; APPLICANT: Decker, Douglas
; APPLICANT: Xiazunhi Li
; TITLE OF INVENTION: Method for Screening for acRAB Transporter Family Inhibitors
; FILE REFERENCE: 6206
; CURRENT APPLICATION NUMBER: US/10/392,234A
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/364,935
; PRIOR FILING DATE: 2002-03-15
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 1048
; TYPE: PRT
; ORGANISM: Salmonella typhimurium
US-10-392-234A-18

```

Query Match	2.0%;	Score 141.5;	DB 6;	Length 1048;
Best Local Similarly	18.1%;	Pred. No. 0.0097;		
Match 145; Conservative	110;	Mismatches 243;	Indels 303;	Gaps 32

```

QY      624  E1NRRTADBLPFAFASYSIYFPIYSIALGSSYSWSKVMDSKRTL--GLGGVAVVVG--A 679
Db      339  EVVKTLVE-----AIIIVFLMWYFLQNF-----RATLLPTIAVPAVVLGGTPA 381
QY      680  VMAANGFESYIGRSSLVITLOVPRFLVSVG--ADNIFIVLEYOGLPRRPGEPREVNIG 737
Db      382  VLAAGF-----SINTLMFG--MVLAIIGLLVDALVVENVERMTBEGLPPEAR 432
QY      738  PALGR-----VAPSMLLCSLSAICFFGLCATPMPAVRTPALTSGIAVLIDFLQMSAFV 792
Db      433  KSMGIOGALVGIAMVLSAFLPMAFFGSGTGAI--YRQFSIIVSAMALSVLVALITLP 490
QY      793  ALLSDSKRQEASRLDVCCEVCKPQELPRPGQEGBLLGFFOKAVYAFBLHWI----- 844
Db      491  ALCA-----TMLKPVAKGDHGBEGKKGFQGFQFNRLFDKSTHHYDSVGNILR 536
QY      845  TRGVLLFLALF-----GVSLYSMCHISVGLDGLALPKDSYLL 884
Db      537  STGRVLLLYLLIVGMAVYFVRLPSSFLPDEDOGVFL--TMVQLPAGATQERQKVLDEVY 595
QY      885  DYFL-----FLNR----- 892

```

```

Db      596 DYILNKEANVESFAVANGPFGPAGRGQNTGIAFVSLKMDADRPGRKKNVEALTQRAITAF 655
Qy      893 -----YEVGAPVYFVTTLTGYNFSSEAGMNAICSSAGCNPSFTQ----- 932
Db      656 SQIKDAMFAENLPALVELGATGATGDFE-----LIDQAGLGHEKLTQARNQLEGEVAK 708
Qy      933 -----KIQVATEPFPQSYLAIPAS-----SMVDFDFDWLTP 963
Db      709 YPDLLVYGRPNGLETPQFKIDIDEXKQALGVSLSDINTTLTGAMGSGYVNDFFD--R 765
Qy      964 SSCRLYISGENKDKFCPCSTVNSLNLCKNCSMTGWSVRPSVEQFKYLPWFELNDRPNIK 102:
Db      766 GRVKKVYVMSEAKYRMLPDDIND-----WYRSGSDQM 798
Qy      1024 CPKGLAAYR-----SVNLTSQGVLAASFMAHKLKNSQDYTEALRARELANANT 107
Db      799 VP--FSAFSSSRMEYGSPLREYNGALPSMELLGOAAQKST--GEANAMMEBLASKUP 852
Qy      1078 ADLRKVPETDPAFEVFPYITINVFEOYLT-----LPEGLFMLSCL----- 1124
Db      853 SGI-----GYDWTGMSYQERLSGQAPALVAISLIVFLCIALALYESMSIPF 899
Qy      1121 -----VPPFAVSCLLGLDLRSGLLNLSIWMILVDYTGFMALMDISYNAVSLN----- 1177
Db      900 SVMLVPLVPGVIGALLAA--TFRGTLNDVYFQVGLTTTIGLSA--KNAILIVFAPADL 952
Qy      1171 -----LVSAVGSVEFVSHITRSPAISTKPTMLBRAKEATISMGSAVFAGVANT 121:
Db      953 MDKSGKGLVEAMLEAVRMRRLPIMLTSLAFMLGVNP-----LVISSG-----AGSGAQ 1004
Qy      1220 NLPGILIV-GLAKAQLOIIFP 1239
Db      1001 NAVGTGVLGAVTATVLAIFP 1021

```

```

RESULT 11
US-10-392-234A-16
; Sequence 16, Application US/10392223AA
; Publication No. US2005025538A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia and Upjohn Corporation
; APPLICANT: Buxser, Steven
; APPLICANT: Poole, Keith
; APPLICANT: Decker, Douglas
; APPLICANT: Kiazhi, Li
; TITLE OF INVENTION: Method for Screening for acarb Transporter Family Inhibitors
; FILE REFERENCE: 6206
; CURRENT APPLICATION NUMBER: US/10/392,234A
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/364,935
; PRIOR FILING DATE: 2002-03-15
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-392-234A-16

```

Query Match	2.0%;	Score 140.5;	DB 6;	Length 1046;
Best Local Similarly	20.5%;	Pred. NO. 0.012;		
Matches 187;	Conservative 126;	Mismatches 305;	Indels 293;	Gaps 49

```

OY 506 NQTLMGQI--SQVBMKHDFLYCANAPLRFKQGTALALSCMDYCAPVPPPLAIGVKGD 563
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 NATIIIGKRIJLTAEQEFENILKVN---PDQSQVLKQVADVG-----Lg---GD 274
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 564 YSEABALIMTESLNNYPA-----GDPRLAQAKWEE--AFLEEMRAFORRMAGMFO 612
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 YS-----INQFNSPASIAIKATGANNALDTKAKIRQTIANLEPFPMQGMKVVPPD 328
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 613 VTFAERGLEDEINNTTAEDLPITATSYIVFLYSLALGYSVSNRVNVDKATL--GL 670
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      329 TTPVVSASH-ERVKTGE-----ALLVELVMTLFLONF-----RATLIPIT 370
Qy      671 GGVAVVIG--AVMAAGFFSYLGISSVLLOVPPVLVSVG--ADNIFILEYORLPR 726
      371 AVPVVLIGFVGLAARGF-----SINTLTWFG--MVLALIGLVDAIVVENVERVMA 421
Qy      727 REG-EREV---HIGALGR--VAPSMILCSLSAICPFIDGALTPMAVRFPALTSLAVI 781
      422 EBGLSPREARAKSMGOIOGALVGIAMVLSVFLPMAFFGGSTGVI--YRQFSTITIVSAMA 479
Qy      782 LDFLLGMSAFVALLSLDSKQGEASRLDVCC---VPRQELPRPGQEGILLGFQKAYAP 838
      480 L-----SYVALI-----LTPALCATMLRPIERGDGHEKKGCFGFNMKMFIS 522
Qy      839 FLHMWTRGVV-----LLEFLALFGVSLYSMCHISVGL-----DQ-----ELALPKD 880
      523 -TTHGVERGVASILTKHAPYLLIYVIVAGMIMFRIPTAPLPDDQGLFAQVQTPPG 581
Qy      881 S-----YLLDYFLFLNRYFEYGADEVYFTTLTGYNPSS-----913
      582 SSAERTQVVVDSMRREVLEK-----ESSSVSVFTVGFNFAGRGQSSGMAFTMLKP 633
Qy      914 -----AGMAI---CSAGCNNSFTOKIOYATEPBGSTLALPASSWDDFID-----959
      634 WEERPGGENSVFELARQAHFFSFYDAMVFA--FAPPSVLELGNATIGDLFLQDQAGVG 691
Qy      960 ----MLTPSSCCCLYISGPNKDFPCSTVA-----SLNCLKKCMST 996
      692 HEVTLQARKKFTMLAQNAPALQVRPNQMSDEPOYKLEIDDEKASALGVSADINSTVSI 751
Qy      997 TWGSV-----RPSVEQFHXYLPWFLNDRPN-----IKCPKGLA--AYST 1034
      752 AMGSSVYNDPDIRGRVYR-----YLQGRPDARMPBDLSKMYVNDKEMVPFNAFAT 805
Qy      1035 -----SVNLTSGQVLASFMAYHKPKRSQDYTEALRARELANITDILAKVGTDP 1088
      806 GKREYSPKTERYNGVPAHEILGEPAGLSSGD--AMAARVE--YVQOLPKGVG--855
Qy      1089 AEFVFPYTTITNVFEQYLT-----LLEBGLFMLSIC-----LVPTFAVS 1127
      856 -----YSWTGLSYERLSSQAPALYALSLVVFCLALYVESMSIPSVMLVVPGLVI 999
Db      1128 CLLGLDLSGGLNLLSTIWMILVDIVGFMALMDISYNAVSLNL-----V 1172
      910 GALLATSMR-GLSNDVFQVGLTTIGLSA-----KNALIVFEAKELHEQKGIYEAL 963
Qy      1173 SAVGMSVERVSHTTRFPAISTKPTWLBRAKEATISGSAVFAVAMTNLPGLIVL-GLAK 1231
      964 EACRMKRLRPIVMTSLAFILGVDP-----LAIISG--AGSGQHAIGTGVIGGMT 1011
Db      1232 AOLIOIFFPRL 1242
Qy      1012 ATVLAIFWVPL 1022

```

```

RESULT 12
US-10-467-657-2648
; Sequence 2648, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04

```

```

; SEQ ID NO 2648
; LENGTH: 1067
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2648

Query Match      1.9%; Score 134; DB 6; Length 1067;
Best Local Similarity 19.4%; Pred. No. 0.039;
Matches 185; Conservative 139; Mismatches 354; Indels 274; Gaps 44;

Qy      452 RHLQVW-SPEAQR--NISLQDICYPALNPDNSTLYVCCINSLLQYQNNRTLLLTANOT 508
      180 RMRIVWDPRKKIQNTNLSPADVGA--LSAQNTQISAGSISGLPAV--RQGT---VTAATV 233
Qy      509 LMGQTSQVDMKQHFLYCANAPLTFKQGTALALSCMADYGAVPVFPALIGYKQDYSEAE 568
      234 AQQQLGTAEFGVILIRANT-----DGSNIYLKDVAKVGLGM-----EDYSSST 277
Qy      569 ALI-----MTSLNNYPAGDPRLOAKLWEAFLEEMRAFORRMAGMFOVTTAERSLE 622
      278 RLNGVNTTGMAWLSN--SGN-AMATAKAVKERLAVLEKYPQGMG--WTPYDTSKFVE 332
Qy      623 DEINRTAADLPFIATSYVIFLYISLALGSYSMSRWVDSKATIGLGSVAIVLAGVMA 682
      333 ISIEKVIHTLIEAMVLVFVMTLFLONT--RTILIFTIVPISL--LGGA-----379
Qy      683 AMGFPSYLGIRSSLVLLOVPPVLVSVDGADNIFIPLYEQRLPRRGEEREVHIGRALGR 742
      380 ---FISYMGH-SINVLTMFMAMILVIGIVDDAIVVENVERIMAGGLPPEKATKKAMGQ 435
Qy      743 VAPSMI-----LCSLSAICFPLGALTMPMAVRTPLTSLGLAVITLFLQMSAFVALLSL 797
      436 ISGAVIGITAVLIVSVFPLMFSGAAGNI--YKQFPLTMASSI-----AFSAFLAL--484
Qy      798 DSKRQASRLDVCCCKPQELPRPGQEGI--LLGFQKAYAP-----LHMWT 845
      485 -----LTPALCATMLKTTPKGHHEKKGCFGMFNKKFDSWTHGYBRVAKVLRKT 536
Qy      846 RGVALLFLALFGVSLYSMCHISVGLDOELAPKDSYLLDYPFLNRYFEVGAUVYFVTT 905
      537 RMNVVYIGLAVGV-----FLFMR-----LPTSLPT 563
Qy      906 LGYNFSSEAGMAIICSSAGC---NNFSFTOKIQVATEPPE--OSYLAIPASSVVD---956
      564 EDQGFV-----MVSQAPAGATKERTDATLAQTOLAKSIPELHNIITVSGSFSGSQNM 619
Qy      957 -----FIDW--LTPSSCCRLYIS-----GPNDKXCPSTVN-----S 986
      620 AMGFALIKDMNERTAGSDAIVAVAGKLTGMMWGTLDGFGISVPPILIELNGSGSLSTN 679
Qy      987 LNCCLKMSITMGSVAPSVQFHKYLPWF-----LNDRENIKCPKGLAAYSTSVN 1037
      680 LQDRNNTGTALLAKKNELIQKRASGLDPSTVRAGGLEDSFQLKIDINRAAAAQGIS 739
Qy      1038 LPSDQVLASRF-MAYHKPLKNSQDYTEALRARELANITADLRK-----VP--1084
      740 FADIRFALASALSSSVSDPFPNGRLQRYVWQADEBARQPADILMLTVPNKSGAVPLS 799
Qy      1085 -----GTD-----PAFEVFPYTTITNVFEQYLT-----1106
      800 TIATVSWENGTEQSVRFNGVPSMKLSASPATGVSTQQAAMAAQKMWDELGGYGFEMGQ 859
Qy      1107 -----LTPBGLFMLSICLV-----PTFAVSCLLLGL-----DLR 1136
      860 SREKAGSGCOTLLIYGLAIVAVLTVLAALYESMSTPLAVIIVPLGLIGAAAGVTRNL 919
Qy      1137 SGLL-NLISIVMLVDIVGFMALMDIS--YNAVSLNLV-----SAVMSVE-----F 1181
      920 EGLGSGVSPFANDIYQVQGVTVWGLSAXKAAIILIFAKDLQKQGSAAVEAALAEARLRF 979
Qy      1182 VSHITRSPA--ISTKPTWLBRAKEATI--SMGSAVAGVAMTNLPGLIVGL 1229
      980 RPLIMTSFAFILGVVPLIYAAGASSASORALIGTIVGMLVGTLLSVFLVPL 1031

```







THIS PAGE BLANK (w8p10)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 23:06:57 ; Search time 188 Seconds  
(without alignments)  
2960.365 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909  
Sequence: 1 MAEAGIRGWLWALLRLA.....GSIKAGAIINFLPNNGROF 1332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/prodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubppaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubppaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6909	100.0	1332	4 US-10-621-758A-4	Sequence 4, Appl1
2	6909	100.0	1332	4 US-10-663-208A-4	Sequence 4, Appl1
3	6909	100.0	1332	4 US-10-646-301A-4	Sequence 4, Appl1
4	6909	100.0	1332	4 US-10-736-769-4	Sequence 4, Appl1
5	6896	99.8	1332	4 US-10-239-316-9	Sequence 9, Appl1
6	6872.5	99.5	1359	4 US-10-621-758A-4	Sequence 44, Appl1
7	6872.5	99.5	1359	4 US-10-663-208A-4	Sequence 44, Appl1
8	6872.5	99.5	1359	4 US-10-646-301A-4	Sequence 44, Appl1
9	6872.5	99.5	1359	4 US-10-736-769-4	Sequence 44, Appl1
10	6536	94.6	1344	5 US-10-450-763-53052	Sequence 53052, A
11	5421.5	78.5	1331	4 US-10-621-758A-2	Sequence 2, Appl1
12	5421.5	78.5	1331	4 US-10-663-208A-2	Sequence 2, Appl1
13	5421.5	78.5	1331	4 US-10-646-301A-2	Sequence 2, Appl1
14	5421.5	78.5	1331	4 US-10-736-769-2	Sequence 12, Appl1
15	5407	78.3	1333	4 US-10-621-758A-12	Sequence 12, Appl1
16	5407	78.3	1333	4 US-10-663-208A-12	Sequence 12, Appl1
17	5407	78.3	1333	4 US-10-646-301A-12	Sequence 12, Appl1
18	5407	78.3	1333	4 US-10-736-769-12	Sequence 12, Appl1
19	4466	64.6	982	5 US-10-450-763-53050	Sequence 53050, A
20	2402.5	34.8	1278	4 US-10-208-731-2	Sequence 2, Appl1
21	2402.5	34.8	1278	4 US-10-741-601-530	Sequence 530, App
22	2402.5	34.8	1278	5 US-10-741-600-1542	Sequence 1542, Ap
23	2402.5	34.8	1278	5 US-10-756-149-4924	Sequence 4924, Ap
24	2385	32.7	1319	4 US-10-208-731-4	Sequence 4, Appl1
25	2262	32.7	1287	6 US-11-097-143-12003	Sequence 12003, A
26	1864.5	27.0	1223	6 US-11-097-143-2679	Sequence 2679, Ap
27	1329	19.2	1170	4 US-10-208-731-6	Sequence 6, Appl1

28	1078	15.6	1296	4 US-10-208-731-9	Sequence 9, Appl1
29	942	13.6	194	4 US-10-239-316-8	Sequence 8, Appl1
30	900.5	13.0	492	4 US-10-424-599-211862	Sequence 211862, Sequence 189288,
31	859.5	12.4	541	4 US-10-424-599-189288	Sequence 19, Appl1
32	651.5	9.4	1447	2 US-08-954-701A-19	Sequence 5, Appl1
33	651.5	9.4	1447	3 US-09-898-533-5	Sequence 5, Appl1
34	651.5	9.4	1447	3 US-09-754-032-19	Sequence 19, Appl1
35	651.5	9.4	1447	4 US-10-421-446-19	Sequence 19, Appl1
36	651.5	9.4	1447	4 US-10-791-844-6	Sequence 6, Appl1
37	646.5	9.4	1434	2 US-08-954-701A-10	Sequence 10, Appl1
38	646.5	9.4	1434	3 US-09-754-032-10	Sequence 10, Appl1
39	646.5	9.4	1434	4 US-10-421-446-10	Sequence 10, Appl1
40	628	9.1	1296	4 US-10-302-279-60	Sequence 60, Appl1
41	598.5	8.7	933	4 US-10-415-934-3	Sequence 3, Appl1
42	597	8.6	821	5 US-10-890-776A-4805	Sequence 4805, Ap
43	588	8.5	954	4 US-10-060-756A-3	Sequence 3, Appl1
44	588	8.5	954	5 US-10-890-776A-3	Sequence 3, Appl1
45	584	8.5	1203	3 US-09-990-046-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1									
US-10-621-758A-4									
; Sequence 4, Application US/10621758A									
; Publication No. US20040093629A1									
; GENERAL INFORMATION:									
; APPLICANT: Altman, Scott W									
; APPLICANT: Wang, Luquan									
; APPLICANT: Graziano, Michael									
; APPLICANT: Margolio, Nick									
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF									
; FILE REFERENCE: JB01603-K-US									
; CURRENT APPLICATION NUMBER: US/10/621,758A									
; CURRENT FILING DATE: 2003-07-17									
; PRIOR APPLICATION NUMBER: 60/397,442									
; PRIOR FILING DATE: 2002-07-19									
; NUMBER OF SEQ ID NOS: 50									
; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO: 4									
; LENGTH: 1332									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-621-758A-4									
Query Match 100.0%; Score 6909; DB 4; Length 1332;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MAEAGIRGWLWALLRLA	OSSEPYTTHHOGYCAPYDEGKPELSGSLMTLSNVCSLSN	60					
DB	1	MAEAGIRGWLWALLRLA	OSSEPYTTHHOGYCAPYDEGKPELSGSLMTLSNVCSLSN	60					
QY	61	TPARKITGHTLILQKICRPLVYPTOACCSAKQVLSIASITKALLTRCPACSDNF	120						
DB	61	TPARKITGHTLILQKICRPLVYPTOACCSAKQVLSIASITKALLTRCPACSDNF	120						
QY	121	VNIHCNNTGSPNOSLFINTRVAQLGAGQLPAVVAEATYQHSFAQOSYDSCRRVPPA	180						
DB	121	VNIHCNNTGSPNOSLFINTRVAQLGAGQLPAVVAEATYQHSFAQOSYDSCRRVPPA	180						
QY	181	ATLAATGTCGVSSALCNARWLNFGDGTNGALPDIPTFHLLPQOAVSGIQPLNEGV	240						
DB	181	ATLAATGTCGVSSALCNARWLNFGDGTNGALPDIPTFHLLPQOAVSGIQPLNEGV	240						
QY	241	ARCNEQSDVAVTSCQDCASCPATARPQALDSTFYLGQMPGSLVLIILCSVPAVVTI	300						
DB	241	ARCNEQSDVAVTSCQDCASCPATARPQALDSTFYLGQMPGSLVLIILCSVPAVVTI	300						
QY	301	LIVGFVAPARADSKVNDPKKGTSLSDKLSFSHTILLGQFQCGWGTWVAWPLTIVLSV	360						
DB	301	LIVGFVAPARADSKVNDPKKGTSLSDKLSFSHTILLGQFQCGWGTWVAWPLTIVLSV	360						

QY 361 IIVVAAAGLVTELTITDPEVLSAPNSQARSEKAFHDOHFGFFFTNOVILTAENRSY 420  
DB 361 IIVVAAAGLVTELTITDPEVLSAPNSQARSEKAFHDOHFGFFFTNOVILTAENRSY 420  
QY 421 RYDSLLGPKNSGILDLLELELELORLRLHLYQWSPBAORNISLODI CYAPLNDPT 480  
DB 421 RYDSLLGPKNSGILDLLELELELORLRLHLYQWSPBAORNISLODI CYAPLNDPT 480  
QY 481 SYDDCINSLLQYFQNNRTLLLLLTANQTLMGQTSQVDMKDHLYCANAPLTFKDGTAAL 540  
DB 481 SYDDCINSLLQYFQNNRTLLLLLTANQTLMGQTSQVDMKDHLYCANAPLTFKDGTAAL 540  
QY 541 SCMDYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNYPADPRLAOKLMEAELEEM 600  
DB 541 SCMDYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNYPADPRLAOKLMEAELEEM 600  
QY 601 RAFORMAGMFOVTFPFAERSLEDEINRTAEDLPFATSYIVIFLYISALGYSMSRV 660  
DB 601 RAFORMAGMFOVTFPFAERSLEDEINRTAEDLPFATSYIVIFLYISALGYSMSRV 660  
QY 661 MYSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSVILQVVPFLVLSGADNIFFLVE 720  
DB 661 MYSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSVILQVVPFLVLSGADNIFFLVE 720  
QY 721 YORLPRRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTEMPAVRTFALISGLAV 780  
DB 721 YORLPRRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTEMPAVRTFALISGLAV 780  
QY 781 IILDFLQMSAFVALSLDSKROBASRLDCCCVKPELPPGGGBGLLGFQKAXAPPL 840  
DB 781 IILDFLQMSAFVALSLDSKROBASRLDCCCVKPELPPGGGBGLLGFQKAXAPPL 840  
QY 841 LHMITRGVLLLFALFGVLSYSMCHI SVGLDQELALPKDSYLLDYFLFMRFEVGAIV 900  
DB 841 LHMITRGVLLLFALFGVLSYSMCHI SVGLDQELALPKDSYLLDYFLFMRFEVGAIV 900  
QY 901 YFVTTLGVNFSSEAGNNAICSSAGCANNFSTQKI QYATEPPEQSYLAIPASSVVDPIWM 960  
DB 901 YFVTTLGVNFSSEAGNNAICSSAGCANNFSTQKI QYATEPPEQSYLAIPASSVVDPIWM 960  
QY 961 LPPSSCRLYISGPNKDKFCPTSVNSLCLKNCKMSITMGSVRPSVDFHXYLPWFLNDR 1020  
DB 961 LPPSSCRLYISGPNKDKFCPTSVNSLCLKNCKMSITMGSVRPSVDFHXYLPWFLNDR 1020  
QY 1021 NIKCPKGLAAYSTSVNLSDQVLA SRFMAVHKPKNSQDYTEALRAARELANITADL 1080  
DB 1021 NIKCPKGLAAYSTSVNLSDQVLA SRFMAVHKPKNSQDYTEALRAARELANITADL 1080  
QY 1081 RAYPGTDPAPFVPPYTTITNVFEQYVLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGL 1140  
DB 1081 RAYPGTDPAPFVPPYTTITNVFEQYVLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGL 1140  
QY 1141 NLLSTMIILVDYMGPMALMDSYNAVSLNLVSAVGMSEFVSHITRSPAIISKPPWLER 1200  
DB 1141 NLLSTMIILVDYMGPMALMDSYNAVSLNLVSAVGMSEFVSHITRSPAIISKPPWLER 1200  
QY 1201 AKEATISMSASAVAGVAMTNLPGLIVLGLAKQLI QIFFRLNLITLLGLHGLVFLV 1260  
DB 1201 AKEATISMSASAVAGVAMTNLPGLIVLGLAKQLI QIFFRLNLITLLGLHGLVFLV 1260  
QY 1261 ILSYGPDPVNPALALEQKRAEBAVAAVMAVASCNHPRSVSTADNIYVNSFEKSGKAG 1320  
DB 1261 ILSYGPDPVNPALALEQKRAEBAVAAVMAVASCNHPRSVSTADNIYVNSFEKSGKAG 1320  
QY 1321 ISNFLPNNGRQF 1332  
DB 1321 ISNFLPNNGRQF 1332

RESULT 2  
US-10-663-208A-4  
; Sequence 4, Application US/10663208A

Publication No. US20040132058A1  
; GENERAL INFORMATION:  
; APPLICANT: Altman, Scott W  
; APPLICANT: Wang, Luquan  
; APPLICANT: Graziano, Michael  
; APPLICANT: Murgolo, Nick  
; TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF  
; FILE REFERENCE: JB01603K2 US  
; CURRENT APPLICATION NUMBER: US/10/663,208A  
; CURRENT FILING DATE: 2003-09-16  
; PRIOR APPLICATION NUMBER: 60/397,442  
; PRIOR FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 10/621,758  
; PRIOR FILING DATE: 2003-07-17  
; PRIOR APPLICATION NUMBER: 10/646,301  
; PRIOR FILING DATE: 2003-08-22  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1332  
; TYPE: PR  
; ORGANISM: Homo sapiens  
US-10-663-208A-4

Query Match 100.0%; Score 6909; DB 4; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEAGLRGWLWALLRLRLAQSPPYTHIQPGYCAFYDECKNPELSGSLMTLSNVCSLSN 60  
DB 1 MAEAGLRGWLWALLRLRLAQSPPYTHIQPGYCAFYDECKNPELSGSLMTLSNVCSLSN 60  
QY 61 TPARKITGDHLLLOKICPRLYTGNTQACSGAKOLVSIASLSTTKALLTRCPACSDNF 120  
DB 61 TPARKITGDHLLLOKICPRLYTGNTQACSGAKOLVSIASLSTTKALLTRCPACSDNF 120  
QY 121 VNLHCNHTCSPNQSLPIINTTRVAOLGAGOLPAVVAVEAFYQHSFPAEQSDSCSRVRVPA 180  
DB 121 VNLHCNHTCSPNQSLPIINTTRVAOLGAGOLPAVVAVEAFYQHSFPAEQSDSCSRVRVPA 180  
QY 181 ATLAVGTMCGVGSALCNARWLNFQDPTGNGLAPLDITFHLLEFGQAVSGSIQPLNEGV 240  
DB 181 ATLAVGTMCGVGSALCNARWLNFQDPTGNGLAPLDITFHLLEFGQAVSGSIQPLNEGV 240  
QY 241 ARCNSQGDVATCSQDCAASCPAIARPOLDSFFYLGOMPGSLVLIILCSVPAVVTI 300  
DB 241 ARCNSQGDVATCSQDCAASCPAIARPOLDSFFYLGOMPGSLVLIILCSVPAVVTI 300  
QY 301 LLVGRVVAAPARKSKVNPBKKTSLSDKLSFSTHTLLGQFPQGWGTWASWPLTILVLSV 360  
DB 301 LLVGRVVAAPARKSKVNPBKKTSLSDKLSFSTHTLLGQFPQGWGTWASWPLTILVLSV 360  
QY 361 IIVVAAAGLVTELTITDPEVLSAPNSQARSEKAFHDOHFGFFFTNOVILTAENRSY 420  
DB 361 IIVVAAAGLVTELTITDPEVLSAPNSQARSEKAFHDOHFGFFFTNOVILTAENRSY 420  
QY 421 RYDSLLGPKNSGILDLLELELELORLRLHLYQWSPBAORNISLODI CYAPLNDPT 480  
DB 421 RYDSLLGPKNSGILDLLELELELORLRLHLYQWSPBAORNISLODI CYAPLNDPT 480  
QY 481 SYDDCINSLLQYFQNNRTLLLLLTANQTLMGQTSQVDMKDHLYCANAPLTFKDGTAAL 540  
DB 481 SYDDCINSLLQYFQNNRTLLLLLTANQTLMGQTSQVDMKDHLYCANAPLTFKDGTAAL 540  
QY 541 SCMDYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNYPADPRLAOKLMEAELEEM 600  
DB 541 SCMDYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNYPADPRLAOKLMEAELEEM 600  
QY 601 RAFORMAGMFOVTFPFAERSLEDEINRTAEDLPFATSYIVIFLYISALGYSMSRV 660  
DB 601 RAFORMAGMFOVTFPFAERSLEDEINRTAEDLPFATSYIVIFLYISALGYSMSRV 660  
QY 661 MYSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSVILQVVPFLVLSGADNIFFLVE 720

|||||  
Db 661 MVDKATLIGGVAVVLGAVMAAMGFFSYGIRSSVLIILQVVPFLVLSVGADNIFIFVLE 720  
Qy 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEALCFPLGALTTPMPAVTFTPLTSGLA 780  
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEALCFPLGALTTPMPAVTFTPLTSGLA 780  
Qy 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPGQEGGLLGFPOKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPGQEGGLLGFPOKAYAPFL 840  
Qy 841 LHMITRGVLLFLALFGVLSYGMCHISVGLDOELAPKDSYLLDYFLFNRYFEVGAPV 900  
Db 841 LHMITRGVLLFLALFGVLSYGMCHISVGLDOELAPKDSYLLDYFLFNRYFEVGAPV 900  
Qy 901 YFVTTTGYNFSSEAGNNAICSSAGCNNSFTQKIQVATEPPEOSYLAIPASSWDDFDW 960  
Db 901 YFVTTTGYNFSSEAGNNAICSSAGCNNSFTQKIQVATEPPEOSYLAIPASSWDDFDW 960  
Qy 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVDFPHKYLPMFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVDFPHKYLPMFLNDRP 1020  
Qy 1021 NIKCPKGLAAVSTSVNLTSDOVLASRFMAVYHKPLKNSODYTEALPAARELANITADL 1080  
Db 1021 NIKCPKGLAAVSTSVNLTSDOVLASRFMAVYHKPLKNSODYTEALPAARELANITADL 1080  
Qy 1081 RKVGTDPAPFEPYITITNVEYQYLTILPEGLFMLSCLVPTFVAVSCLLGLDLSGGL 1140  
Db 1081 RKVGTDPAPFEPYITITNVEYQYLTILPEGLFMLSCLVPTFVAVSCLLGLDLSGGL 1140  
Qy 1141 NLSIYMIIVDYGFMAIMDISYNAVSLINLVSAGSVFVSHITRSPAISTRKPTWLER 1200  
Db 1141 NLSIYMIIVDYGFMAIMDISYNAVSLINLVSAGSVFVSHITRSPAISTRKPTWLER 1200  
Qy 1201 AKKATISMSGSAFVAGVAMTLPGLIVLGLAKAOLIOIFFRLMLITLLGLLGLVFLV 1260  
Db 1201 AKKATISMSGSAFVAGVAMTLPGLIVLGLAKAOLIOIFFRLMLITLLGLLGLVFLV 1260  
Qy 1261 ILSYVGPDPVAPALAEQKRAEBAVAAVMAVASCNHPBSRVSTADNITYNHSFEKIGAGA 1320  
Db 1261 ILSYVGPDPVAPALAEQKRAEBAVAAVMAVASCNHPBSRVSTADNITYNHSFEKIGAGA 1320  
Qy 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332

RESULT 3  
US-10-646-301A-4  
: Sequence 4, Application US/10646301A  
: Publication No. US20040137467A1  
: GENERAL- INFORMATION:  
: APPLICANT: Altman, Scott W  
: APPLICANT: Wang, Luquan  
: APPLICANT: Graziano, Michael  
: APPLICANT: Margolo, Nick  
: TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF  
: FILE REFERENCE: J01603-K1-US  
: CURRENT APPLICATION NUMBER: US/10/646,301A  
: PRIOR FILING DATE: 2003-08-22  
: PRIOR APPLICATION NUMBER: 60/397,442  
: PRIOR FILING DATE: 2002-07-19  
: PRIOR APPLICATION NUMBER: 10/621,758  
: PRIOR FILING DATE: 2003-07-17  
: NUMBER OF SEQ ID NOS: 50  
: SOFTWARE: Patent version 3.1  
: SEQ ID NO 4  
: LENGTH: 1332  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-10-646-301A-4

Query Match 100.0%; Score 6909; DB 4; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MAEAGIRGMLMALRLAQLQSEPTTTHQPGYCAFDEGCKNPELSGSLMTLSNVCLSN 60  
Db 1 MAEAGIRGMLMALRLAQLQSEPTTTHQPGYCAFDEGCKNPELSGSLMTLSNVCLSN 60  
Qy 61 TPAARKITGHLILLOKICRPLYTGPMTOACCSAKQVSLSEASTITKALLTRCPACSDNF 120  
Db 61 TPAARKITGHLILLOKICRPLYTGPMTOACCSAKQVSLSEASTITKALLTRCPACSDNF 120  
Qy 121 VNLHCNTPSNOSLFTINTVVAQLGAGQLPAVVAEAFYQHSFAQSDYDSCSRVRVPA 180  
Db 121 VNLHCNTPSNOSLFTINTVVAQLGAGQLPAVVAEAFYQHSFAQSDYDSCSRVRVPA 180  
Qy 181 ATLAAGTCMGVGSALCNARWLNFGQDTGNGIAPADITFHLLEPCQAVGSGIQPLNEGV 240  
Db 181 ATLAAGTCMGVGSALCNARWLNFGQDTGNGIAPADITFHLLEPCQAVGSGIQPLNEGV 240  
Qy 241 ARCNEQGDVATCSQDCAASCPALARPQALDSTYLGQMPGSLVLIILCSVPAVVTI 300  
Db 241 ARCNEQGDVATCSQDCAASCPALARPQALDSTYLGQMPGSLVLIILCSVPAVVTI 300  
Qy 301 ILVGFVAPARBDKSNVDPKGTSLSDKLSFTHTLLGQFQSGMTVAVSMPTIIVLSV 360  
Db 301 ILVGFVAPARBDKSNVDPKGTSLSDKLSFTHTLLGQFQSGMTVAVSMPTIIVLSV 360  
Qy 361 IPVALAAGLVFTELTTDPVELMSAPNSQARSEKAFHDOFGBFRTNOVILTPARRSSY 420  
Db 361 IPVALAAGLVFTELTTDPVELMSAPNSQARSEKAFHDOFGBFRTNOVILTPARRSSY 420  
Qy 421 RYDSSLGPRNFGIIDLLELLELQERLRLQVWSPBAQNSISLODICTAPLNPDT 480  
Db 421 RYDSSLGPRNFGIIDLLELLELQERLRLQVWSPBAQNSISLODICTAPLNPDT 480  
Qy 481 SLVDCINSLLOVFOUNRTLLLTANOTLMGQSOVDMQDHPLYCANAPLTFDGTALAL 540  
Db 481 SLVDCINSLLOVFOUNRTLLLTANOTLMGQSOVDMQDHPLYCANAPLTFDGTALAL 540  
Qy 541 SCNADYGAPVFPPLAIGYKGDYSEAEALIMFSLNNYVADGPRLOAKLMEAEFLSEM 600  
Db 541 SCNADYGAPVFPPLAIGYKGDYSEAEALIMFSLNNYVADGPRLOAKLMEAEFLSEM 600  
Qy 601 RAFORMAGNFQYTFARSLEDEINRTAEDLPFATSYIVFLYISLALGYSMSRV 660  
Db 601 RAFORMAGNFQYTFARSLEDEINRTAEDLPFATSYIVFLYISLALGYSMSRV 660  
Qy 661 MVDKATLIGGVAVVLGAVMAAMGFFSYGIRSSVLIILQVVPFLVLSVGADNIFIFVLE 720  
Db 661 MVDKATLIGGVAVVLGAVMAAMGFFSYGIRSSVLIILQVVPFLVLSVGADNIFIFVLE 720  
Qy 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEALCFPLGALTTPMPAVTFTPLTSGLA 780  
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEALCFPLGALTTPMPAVTFTPLTSGLA 780  
Qy 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPGQEGGLLGFPOKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPGQEGGLLGFPOKAYAPFL 840  
Qy 841 LHMITRGVLLFLALFGVLSYGMCHISVGLDOELAPKDSYLLDYFLFNRYFEVGAPV 900  
Db 841 LHMITRGVLLFLALFGVLSYGMCHISVGLDOELAPKDSYLLDYFLFNRYFEVGAPV 900  
Qy 901 YFVTTTGYNFSSEAGNNAICSSAGCNNSFTQKIQVATEPPEOSYLAIPASSWDDFDW 960  
Db 901 YFVTTTGYNFSSEAGNNAICSSAGCNNSFTQKIQVATEPPEOSYLAIPASSWDDFDW 960  
Qy 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVDFPHKYLPMFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVDFPHKYLPMFLNDRP 1020  
Qy 1021 NIKCPKGLAAVSTSVNLTSDOVLASRFMAVYHKPLKNSODYTEALPAARELANITADL 1080

Db 1021 NIKCPKGGIAAYSTVNLTSDDQVLA SRPMAVHKPLKNSQDYTEARARELANITADL 1080  
Qy 1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCULLGLDLSGGL 1140  
Db 1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCULLGLDLSGGL 1140  
Qy 1141 NILSTIVMLIVDTVGFMALMDISYNVSLINLVSAVMSVEFVSHITRSFAISTKPTWLER 1200  
Db 1141 NILSTIVMLIVDTVGFMALMDISYNVSLINLVSAVMSVEFVSHITRSFAISTKPTWLER 1200  
Qy 1201 AKEATISMGSAVFAVAMNTNLPGLIVLGLAKAQLIQIFPRLNLITLLGLHGVFLPV 1260  
Db 1201 AKEATISMGSAVFAVAMNTNLPGLIVLGLAKAQLIQIFPRLNLITLLGLHGVFLPV 1260  
Qy 1261 ILSYVGPVNPALALEQKRAEBAVAVVAASCPNHSRSTADNIYVNHSPFGS1KGAGA 1320  
Db 1261 ILSYVGPVNPALALEQKRAEBAVAVVAASCPNHSRSTADNIYVNHSPFGS1KGAGA 1320  
Qy 1321 ISNPLPNNGRQF 1332  
Db 1321 ISNPLPNNGRQF 1332

RESULT 4  
US-10-736-769-4  
Sequence 4, Application US/10736769  
Publication No. US20040161838A1  
GENERAL INFORMATION:  
APPLICANT: Altman, Scott W  
APPLICANT: Wang, Luquan  
APPLICANT: Graziano, Michael  
APPLICANT: Murgolo, Nick  
TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF  
FILE REFERENCE: JB01603-K3-US  
CURRENT APPLICATION NUMBER: US/10/736,769  
CURRENT FILING DATE: 2003-12-16  
PRIOR APPLICATION NUMBER: 60/397,442  
PRIOR FILING DATE: 2002-07-19  
PRIOR APPLICATION NUMBER: 10/621,758  
PRIOR FILING DATE: 2003-07-17  
PRIOR APPLICATION NUMBER: 10/646,301  
PRIOR FILING DATE: 2003-08-22  
PRIOR APPLICATION NUMBER: 10/663,208  
PRIOR FILING DATE: 2003-09-16  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 4  
LENGTH: 1332  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-736-769-4

Query Match 100.0%; Score 6309; DB 4; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MAEAGRGWLMALRLRLAQSEBYTTIHQPGYCAFYDECGKPELSGMTLSNVCLSN 60  
Db 1 MAEAGRGWLMALRLRLAQSEBYTTIHQPGYCAFYDECGKPELSGMTLSNVCLSN 60  
Qy 61 TPARKITGDHLITLQKICPRLYTGPNTQACCSAKQVSLBASISITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLITLQKICPRLYTGPNTQACCSAKQVSLBASISITKALLTRCPACSDNF 120  
Qy 121 VNLHCNNTSPNOSLFINTRVAQLGAGQLPANNVVEAFYQHSFAQSYDSCSRVVPAA 180  
Db 121 VNLHCNNTSPNOSLFINTRVAQLGAGQLPANNVVEAFYQHSFAQSYDSCSRVVPAA 180  
Qy 181 ATLAVGTMGVYGSALCAKQRMNLNFOGDTGNGLAAPDITFHLLEPQAVGSGIQPLNEGV 240  
Db 181 ATLAVGTMGVYGSALCAKQRMNLNFOGDTGNGLAAPDITFHLLEPQAVGSGIQPLNEGV 240

Qy 241 AHCNBSQGDVATCCSCOPCAACPAIARPOLDSTFFYIGQMPGSLVLIILCSVAVVTI 300  
Db 241 AHCNBSQGDVATCCSCOPCAACPAIARPOLDSTFFYIGQMPGSLVLIILCSVAVVTI 300  
Qy 301 LTVGFRVAPARDKSMVPPKGTSLSDLSFSTHTLLAQFOFGMGWTWASWPLTILVLSV 360  
Db 301 LTVGFRVAPARDKSMVPPKGTSLSDLSFSTHTLLAQFOFGMGWTWASWPLTILVLSV 360  
Qy 361 IPVVALAGLVFTELTTDPVELMSAPNSQARSEKAFHQHFQPFRTNQVILTAENRSY 420  
Db 361 IPVVALAGLVFTELTTDPVELMSAPNSQARSEKAFHQHFQPFRTNQVILTAENRSY 420  
Qy 421 RYDSILLGPKNFGSLDLDLLELELOERLPHLOVSPBAQRNLSLODICAAPLNPT 480  
Db 421 RYDSILLGPKNFGSLDLDLLELELOERLPHLOVSPBAQRNLSLODICAAPLNPT 480  
Qy 481 SLYDCCINSLOYPQNNRTLLLTANQTLMGQTSQVMDKHFLYCANAPLTFKQDTALAL 540  
Db 481 SLYDCCINSLOYPQNNRTLLLTANQTLMGQTSQVMDKHFLYCANAPLTFKQDTALAL 540  
Qy 541 SCMDYGAVPFPFLAIGGYKGDYSEAEALIMTSLANNYPAGDPRLAQAKLMEBAFLEBM 600  
Db 541 SCMDYGAVPFPFLAIGGYKGDYSEAEALIMTSLANNYPAGDPRLAQAKLMEBAFLEBM 600  
Qy 601 RAFORRMAGMFOVYTTAERSLEDEINRTTAEDLP1FATS1YIFLYISIALGSSWSRV 660  
Db 601 RAFORRMAGMFOVYTTAERSLEDEINRTTAEDLP1FATS1YIFLYISIALGSSWSRV 660  
Qy 661 MVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSLVILQVVPFLVSGADNIFVYLE 720  
Db 661 MVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSLVILQVVPFLVSGADNIFVYLE 720  
Qy 721 YQRLRRRPEPREVHIGRALGRVAPSMILCSISEAICEFLGALTMPAVRTFALTSGLA 780  
Db 721 YQRLRRRPEPREVHIGRALGRVAPSMILCSISEAICEFLGALTMPAVRTFALTSGLA 780  
Qy 781 IIDFLQMSAFALLSLDSKQEARLDVCCCVKQELPPRQSGGLLGFQKAYAPL 840  
Db 781 IIDFLQMSAFALLSLDSKQEARLDVCCCVKQELPPRQSGGLLGFQKAYAPL 840  
Qy 841 LHMITRGVLLFLALPGVSLYSMSCHISVGLDQELALPKDSYLABFLNRYFEVAGPV 900  
Db 841 LHMITRGVLLFLALPGVSLYSMSCHISVGLDQELALPKDSYLABFLNRYFEVAGPV 900  
Qy 901 YFVTLTGYNFSSEAGMNAICSSAGCNNSFTQKIYATEPBPQSYLAI PASSWVDFTDW 960  
Db 901 YFVTLTGYNFSSEAGMNAICSSAGCNNSFTQKIYATEPBPQSYLAI PASSWVDFTDW 960  
Qy 961 LTPSSCCRLYISGPKDKFCPSTVNSLNCIKKCMSITNGSVRPSYEQFHKYLPWFLNDRP 1020  
Db 961 LTPSSCCRLYISGPKDKFCPSTVNSLNCIKKCMSITNGSVRPSYEQFHKYLPWFLNDRP 1020  
Qy 1021 NIKCPKGGIAAYSTVNLTSDDQVLA SRPMAVHKPLKNSQDYTEARARELANITADL 1080  
Db 1021 NIKCPKGGIAAYSTVNLTSDDQVLA SRPMAVHKPLKNSQDYTEARARELANITADL 1080  
Qy 1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCULLGLDLSGGL 1140  
Db 1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCULLGLDLSGGL 1140  
Qy 1141 NILSTIVMLIVDTVGFMALMDISYNVSLINLVSAVMSVEFVSHITRSFAISTKPTWLER 1200  
Db 1141 NILSTIVMLIVDTVGFMALMDISYNVSLINLVSAVMSVEFVSHITRSFAISTKPTWLER 1200  
Qy 1201 AKEATISMGSAVFAVAMNTNLPGLIVLGLAKAQLIQIFPRLNLITLLGLHGVFLPV 1260  
Db 1201 AKEATISMGSAVFAVAMNTNLPGLIVLGLAKAQLIQIFPRLNLITLLGLHGVFLPV 1260  
Qy 1261 ILSYVGPVNPALALEQKRAEBAVAVVAASCPNHSRSTADNIYVNHSPFGS1KGAGA 1320  
Db 1261 ILSYVGPVNPALALEQKRAEBAVAVVAASCPNHSRSTADNIYVNHSPFGS1KGAGA 1320  
Qy 1321 ISNPLPNNGRQF 1332



```
Db      1321  ISNFLPNNGRPF 1332
|||||
RESULT 5
US-10-239-316-9
; Sequence 9, Application US/10239316
; Publication No. US20030125253A1
; GENERAL INFORMATION:
; APPLICANT: TANIYAMA, Yoshio
; APPLICANT: KITA, Shunbun
; APPLICANT: SATOMI, Tomoko Komiyaama
; TITLE OF INVENTION: No. US20030125253A1e1 Protein, Process for Producing The Same And
; FILE REFERENCE: 2703USOP
; CURRENT APPLICATION NUMBER: US/10/239, 316
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: PCT/JP01/02279
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: JP2000-088595
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 59
; SEQ ID NO 9
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Human
US-10-239-316-9

Query Match      99.8%; Score 6896; DB 4; Length 1332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  MABAGLRGWLIMALLRLAASEPYTTIHOPGYCAFYDECGNPELSGLMTLSNVSCLSN 60
Db      1  MABAGLRGWLIMALLRLAASEPYTTIHOPGYCAFYDECGNPELSGLMTLSNVSCLSN 60
Qy      61  TPARKITGDHILLOKICPRLTYGPNTOACCSAQVLSLEASLITTAALTTRCPACSDNF 120
Db      61  TPARKITGDHILLOKICPRLTYGPNTOACCSAQVLSLEASLITTAALTTRCPACSDNF 120
Qy      121  VNHHCHNTGSPNOSLFINTRVNQLGAGQLPAVVAYEAFYOHSAEBSYDSCSVRVPA 180
Db      121  VNHHCHNTGSPNOSLFINTRVNQLGAGQLPAVVAYEAFYOHSAEBSYDSCSVRVPA 180
Qy      121  VNHHCHNTGSPNOSLFINTRVNQLGAGQLPAVVAYEAFYOHSAEBSYDSCSVRVPA 180
Db      121  VNHHCHNTGSPNOSLFINTRVNQLGAGQLPAVVAYEAFYOHSAEBSYDSCSVRVPA 180
Qy      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGQAVSGIQPLNEGV 240
Db      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGQAVSGIQPLNEGV 240
Qy      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGQAVSGIQPLNEGV 240
Db      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGQAVSGIQPLNEGV 240
Qy      241  ARCNEGQDDVATSCODCAASCPALARPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300
Db      241  ARCNEGQDDVATSCODCAASCPALARPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300
Qy      301  LLYGFRVAPARDKSKMDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
Db      301  LLYGFRVAPARDKSKMDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
Qy      301  LLYGFRVAPARDKSKMDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
Db      301  LLYGFRVAPARDKSKMDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
Qy      361  IPVVALAAGLYFTLTTPVEYELMSAPNSQARSEKAFHDQHGPFPRINOVILTPNRSY 420
Db      361  IPVVALAAGLYFTLTTPVEYELMSAPNSQARSEKAFHDQHGPFPRINOVILTPNRSY 420
Qy      421  RYDSSLGPKNFSGILDLLELLELLEQLRLHLQVNSPEAQRNLSIODICYAPLNDPT 480
Db      421  RYDSSLGPKNFSGILDLLELLELLEQLRLHLQVNSPEAQRNLSIODICYAPLNDPT 480
Qy      421  RYDSSLGPKNFSGILDLLELLELLEQLRLHLQVNSPEAQRNLSIODICYAPLNDPT 480
Db      421  RYDSSLGPKNFSGILDLLELLELLEQLRLHLQVNSPEAQRNLSIODICYAPLNDPT 480
Qy      481  SLVDCCNSLLQYQNNRTLLLTANOTLMQTSQVDMKHFLYCANAPLTFKGTALAL 540
Db      481  SLVDCCNSLLQYQNNRTLLLTANOTLMQTSQVDMKHFLYCANAPLTFKGTALAL 540
Qy      481  SLVDCCNSLLQYQNNRTLLLTANOTLMQTSQVDMKHFLYCANAPLTFKGTALAL 540
Db      481  SLVDCCNSLLQYQNNRTLLLTANOTLMQTSQVDMKHFLYCANAPLTFKGTALAL 540
Qy      541  SCNADYAPVPPFLAIGGYKGYSEAEALMTFSLNNTYPAQDRLAOKLMEAPLEEM 600
Db      541  SCNADYAPVPPFLAIGGYKGYSEAEALMTFSLNNTYPAQDRLAOKLMEAPLEEM 600
Qy      541  SCNADYAPVPPFLAIGGYKGYSEAEALMTFSLNNTYPAQDRLAOKLMEAPLEEM 600
Db      541  SCNADYAPVPPFLAIGGYKGYSEAEALMTFSLNNTYPAQDRLAOKLMEAPLEEM 600
Qy      601  RAFORMAGMFQVTFARSLDEINRTTADLPITATSYVIFLYISLAGSYSSMSRV 660
Db      601  RAFORMAGMFQVTFARSLDEINRTTADLPITATSYVIFLYISLAGSYSSMSRV 660
```

```
Db      601  RAFORMAGMFQVTFARSLDEINRTTADLPITATSYVIFLYISLAGSYSSMSRV 660
Qy      661  MVDKATILGLGVAVVLGVAAMAAGFPSSYLGIRSSLVILQVPELVLSVAGDNIFILE 720
Db      661  MVDKATILGLGVAVVLGVAAMAAGFPSSYLGIRSSLVILQVPELVLSVAGDNIFILE 720
Qy      721  YORLPRRGGPRRHHIGRALGRVAPSWLCSLSSEALTFEFGALTTPMAVNTFALTSLAV 780
Db      721  YORLPRRGGPRRHHIGRALGRVAPSWLCSLSSEALTFEFGALTTPMAVNTFALTSLAV 780
Qy      781  ILDFLLQMSAFVALLSLDSKROEASRLDVCCCVKPOELPPGQEGILLGFFQKAYAPFL 840
Db      781  ILDFLLQMSAFVALLSLDSKROEASRLDVCCCVKPOELPPGQEGILLGFFQKAYAPFL 840
Qy      841  LHMITRGVALLFLALFGVSLYSWCHISVGLDELAPKDSYLLDYELFLNRYEAGAPV 900
Db      841  LHMITRGVALLFLALFGVSLYSWCHISVGLDELAPKDSYLLDYELFLNRYEAGAPV 900
Qy      901  YFVTTLAGYNFSSBAGMAITCSAGCNNSFTQKIQVATEPEQSYLAIPASSWVDPTIDW 960
Db      901  YFVTTLAGYNFSSBAGMAITCSAGCNNSFTQKIQVATEPEQSYLAIPASSWVDPTIDW 960
Qy      961  LTPSSCCRLYISGPNKDKPCPTVNSLNCXKMSITMGSVRSVBEQFHXYLPFLNDRP 1020
Db      961  LTPSSCCRLYISGPNKDKPCPTVNSLNCXKMSITMGSVRSVBEQFHXYLPFLNDRP 1020
Qy      1021  NIKCPKGLAAYSTSVNLTSQGVLASRFMAHYHKLNSODYTEALPAAREBLANITADL 1080
Db      1021  NIKCPKGLAAYSTSVNLTSQGVLASRFMAHYHKLNSODYTEALPAAREBLANITADL 1080
Qy      1081  RKVPGTDPAFEVPEYTTITNVFYEQYLTILBEGLPMLSCLVPTFPAVSCLLGLDLRSGLL 1140
Db      1081  RKVPGTDPAFEVPEYTTITNVFYEQYLTILBEGLPMLSCLVPTFPAVSCLLGLDLRSGLL 1140
Qy      1141  NLSTVWILVDYGFALMDISYNAVSLINUNAVGMSVRFVSHITSFALSTKPTWLER 1200
Db      1141  NLSTVWILVDYGFALMDISYNAVSLINUNAVGMSVRFVSHITSFALSTKPTWLER 1200
Qy      1201  AKERTISGSAVFAGVAMTILPGILVYGLAKAOLIQIFFRLNLLITLGLHGLVPLPV 1260
Db      1201  AKERTISGSAVFAGVAMTILPGILVYGLAKAOLIQIFFRLNLLITLGLHGLVPLPV 1260
Qy      1261  ILSTYGPDVNPALALEQKRAEBAVAAMVASCPNHPRSVSTADNIYVNSFEBSIKGAGA 1320
Db      1261  ILSTYGPDVNPALALEQKRAEBAVAAMVASCPNHPRSVSTADNIYVNSFEBSIKGAGA 1320
Qy      1321  ISNFLPNNGRPF 1332
Db      1321  ISNFLPNNGRPF 1332

RESULT 6
US-10-621-758A-44
; Sequence 44, Application US/10621758A
; Publication No. US20040093629A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-K-US
; CURRENT APPLICATION NUMBER: US/10/621,758A
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/397,442
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 1359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-621-758A-44
```

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;  
 Best Local Similarity 97.9%; Pred. No. 0;  
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY 1 M E A G I R G W L M A L L I R L A O S E P Y T T I H O P G Y C A F Y D E C G K N P E L S G S I M T L S N V S C L S N 60  
 1 M A E A G I R G W L M A L L I R L A O S E P Y T T I H O P G Y C A F Y D E C G K N P E L S G S I M T L S N V S C L S N 60  
 DB 1 M A E A G I R G W L M A L L I R L A O S E P Y T T I H O P G Y C A F Y D E C G K N P E L S G S I M T L S N V S C L S N 60

QY 61 T P A R K I T G D H L I L L O K I C P R L Y T G P N T O A C S A K O L V S L E A S I S T K A L L T R C P A C S D N F 120  
 61 T P A R K I T G D H L I L L O K I C P R L Y T G P N T O A C S A K O L V S L E A S I S T K A L L T R C P A C S D N F 120  
 DB 61 T P A R K I T G D H L I L L O K I C P R L Y T G P N T O A C S A K O L V S L E A S I S T K A L L T R C P A C S D N F 120

QY 121 V N L H C H N T S P N O S L F I N T R V A Q L G A G O L P A V V A Y E A F Y Q H S F A Q S Y D S C S R V P V P A 180  
 121 V N L H C H N T S P N O S L F I N T R V A Q L G A G O L P A V V A Y E A F Y Q H S F A Q S Y D S C S R V P V P A 180  
 DB 121 V N L H C H N T S P N O S L F I N T R V A Q L G A G O L P A V V A Y E A F Y Q H S F A Q S Y D S C S R V P V P A 180

QY 181 A T L A V G T M C G Y G S A L C N A Q R W L N F O G D T G N G L A P D I T F H L L E P Q A V G S G I O P L N E G V 240  
 181 A T L A V G T M C G Y G S A L C N A Q R W L N F O G D T G N G L A P D I T F H L L E P Q A V G S G I O P L N E G V 240  
 DB 181 A T L A V G T M C G Y G S A L C N A Q R W L N F O G D T G N G L A P D I T F H L L E P Q A V G S G I O P L N E G V 240

QY 241 A R C N E S Q G D V A T C S C O D C A A S C P A I A R P Q A L D S T F Y L G M P G S V L I I I L C S V F A V V T I 300  
 241 A R C N E S Q G D V A T C S C O D C A A S C P A I A R P Q A L D S T F Y L G M P G S V L I I I L C S V F A V V T I 300  
 DB 241 A R C N E S Q G D V A T C S C O D C A A S C P A I A R P Q A L D S T F Y L G M P G S V L I I I L C S V F A V V T I 300

QY 301 L I V G F R V A P A R D K S K N V D P K K G T S L S D K L S F S T H T I L G O F P O G M G T W A S W P L T I I V L S V 360  
 301 L I V G F R V A P A R D K S K N V D P K K G T S L S D K L S F S T H T I L G O F P O G M G T W A S W P L T I I V L S V 360  
 DB 301 L I V G F R V A P A R D K S K N V D P K K G T S L S D K L S F S T H T I L G O F P O G M G T W A S W P L T I I V L S V 360

QY 361 I P V A L A A G L V F E L T T D P V E L M S A P N S O A R S E K A F H D O F G F F F T N O V I I L A P R S S Y 420  
 361 I P V A L A A G L V F E L T T D P V E L M S A P N S O A R S E K A F H D O F G F F F T N O V I I L A P R S S Y 420  
 DB 361 I P V A L A A G L V F E L T T D P V E L M S A P N S O A R S E K A F H D O F G F F F T N O V I I L A P R S S Y 420

QY 421 R Y D S L L G R K N F S G I L D L D L L E L L E L O E R L R H L Q V S P E A O R N I S L O D I C V A P L N P D N T 480  
 421 R Y D S L L G R K N F S G I L D L D L L E L L E L O E R L R H L Q V S P E A O R N I S L O D I C V A P L N P D N T 480  
 DB 421 R Y D S L L G R K N F S G I L D L D L L E L L E L O E R L R H L Q V S P E A O R N I S L O D I C V A P L N P D N T 480

QY 481 S L Y D C C I N S L L O Y F O N N R T L L L I T A N Q T L M G O T S O V D M K D H F L Y C A N A P L T F E O G T A L A L 540  
 481 S L Y D C C I N S L L O Y F O N N R T L L L I T A N Q T L M G O T S O V D M K D H F L Y C A N A P L T F E O G T A L A L 540  
 DB 481 S L Y D C C I N S L L O Y F O N N R T L L L I T A N Q T L M G O T S O V D M K D H F L Y C A N A P L T F E O G T A L A L 540

QY 541 S C M A D V G A V F P P L A I G Y K G K O Y S E A E A L I M F S L N N Y P A G O B R L A O K L M E A F L E M 600  
 541 S C M A D V G A V F P P L A I G Y K G K O Y S E A E A L I M F S L N N Y P A G O B R L A O K L M E A F L E M 600  
 DB 541 S C M A D V G A V F P P L A I G Y K G K O Y S E A E A L I M F S L N N Y P A G O B R L A O K L M E A F L E M 600

QY 601 R A F O R R M A G M F O Y T F T A E R S L E D E I N R T T A E D L P I F A T S Y I V F L Y I S L A G S Y S S M S R V 660  
 601 R A F O R R M A G M F O Y T F T A E R S L E D E I N R T T A E D L P I F A T S Y I V F L Y I S L A G S Y S S M S R V 660  
 DB 601 R A F O R R M A G M F O Y T F T A E R S L E D E I N R T T A E D L P I F A T S Y I V F L Y I S L A G S Y S S M S R V 660

QY 661 M Y D S K A T T L G I G V A V V L G A V M A M G F F S Y I G I R S S L V I I L Q V P F L V L S V G A D N I F I F V L E 720  
 661 M Y D S K A T T L G I G V A V V L G A V M A M G F F S Y I G I R S S L V I I L Q V P F L V L S V G A D N I F I F V L E 720  
 DB 661 M Y D S K A T T L G I G V A V V L G A V M A M G F F S Y I G I R S S L V I I L Q V P F L V L S V G A D N I F I F V L E 720

QY 721 Y O R L P R P G B P R P V H I G R A L G R V A P S M L C S L S E A L C F P L G A L T P M P A R T F A L T S G L A V 780  
 721 Y O R L P R P G B P R P V H I G R A L G R V A P S M L C S L S E A L C F P L G A L T P M P A R T F A L T S G L A V 780  
 DB 721 Y O R L P R P G B P R P V H I G R A L G R V A P S M L C S L S E A L C F P L G A L T P M P A R T F A L T S G L A V 780

QY 781 I I D F L L O M S A F V A L L S L D S K R O E A S R L D V C C V K P O E L P P P G O G E G L L G F P O K A Y A P L 840  
 781 I I D F L L O M S A F V A L L S L D S K R O E A S R L D V C C V K P O E L P P P G O G E G L L G F P O K A Y A P L 840  
 DB 781 I I D F L L O M S A F V A L L S L D S K R O E A S R L D V C C V K P O E L P P P G O G E G L L G F P O K A Y A P L 840

QY 841 L H W I T R G V V L L F L A L F G V S L Y S M C H I S V G L D O E L A L P K D S Y L D Y F L P L A N R Y F E V G A P V 900  
 841 L H W I T R G V V L L F L A L F G V S L Y S M C H I S V G L D O E L A L P K D S Y L D Y F L P L A N R Y F E V G A P V 900  
 DB 841 L H W I T R G V V L L F L A L F G V S L Y S M C H I S V G L D O E L A L P K D S Y L D Y F L P L A N R Y F E V G A P V 900

QY 901 Y F P T T L G V N S S A G N N A I C S S A G C N N S F T O K I Q V A T E P P E O S Y A I P A S S A V D P F I M 960  
 901 Y F P T T L G V N S S A G N N A I C S S A G C N N S F T O K I Q V A T E P P E O S Y A I P A S S A V D P F I M 960  
 DB 901 Y F P T T L G V N S S A G N N A I C S S A G C N N S F T O K I Q V A T E P P E O S Y A I P A S S A V D P F I M 960

QY 961 L T P S S C R L Y I S G P N K D K C P S T V N S L N C L K N C M S I T M G S V R P S V E Q F H K Y L P M F L N D R P 1020  
 961 L T P S S C R L Y I S G P N K D K C P S T V N S L N C L K N C M S I T M G S V R P S V E Q F H K Y L P M F L N D R P 1020  
 DB 961 L T P S S C R L Y I S G P N K D K C P S T V N S L N C L K N C M S I T M G S V R P S V E Q F H K Y L P M F L N D R P 1020

QY 1021 N I K C P K G L A A S T S V N L T S D G O V L ----- A S R F M A Y H 1053  
 1021 N I K C P K G L A A S T S V N L T S D G O V L D Y A I L S P R L E Y S G T I S A H C N L Y L D S A S R F M A Y H 1080  
 DB 1021 N I K C P K G L A A S T S V N L T S D G O V L D Y A I L S P R L E Y S G T I S A H C N L Y L D S A S R F M A Y H 1080

QY 1054 K P L K S O D Y T E A L R A A R E L A N I T A D L R K V P G T D P A F E V P Y T I T N V F E Q Y L T I L P E G L 1113  
 1054 K P L K S O D Y T E A L R A A R E L A N I T A D L R K V P G T D P A F E V P Y T I T N V F E Q Y L T I L P E G L 1113  
 DB 1054 K P L K S O D Y T E A L R A A R E L A N I T A D L R K V P G T D P A F E V P Y T I T N V F E Q Y L T I L P E G L 1113

QY 1114 F M L S I C L V P T F A V S C L L G L D R S G L N L S I V M I L V D T V G F M A L M D I S Y N A V S L I N L V S 1173  
 1114 F M L S I C L V P T F A V S C L L G L D R S G L N L S I V M I L V D T V G F M A L M D I S Y N A V S L I N L V S 1173  
 DB 1114 F M L S I C L V P T F A V S C L L G L D R S G L N L S I V M I L V D T V G F M A L M D I S Y N A V S L I N L V S 1173

QY 1174 A Y C M S V E F S H I T R S P A I S T K P T W L E R A K E A T I S M G S A F A G V A M T N L P G I I V L G L A K O 1233  
 1174 A Y C M S V E F S H I T R S P A I S T K P T W L E R A K E A T I S M G S A F A G V A M T N L P G I I V L G L A K O 1233  
 DB 1174 A Y C M S V E F S H I T R S P A I S T K P T W L E R A K E A T I S M G S A F A G V A M T N L P G I I V L G L A K O 1233

QY 1234 L I Q I F F F R L N L I T L L G L H G L V F L P V I L S Y G P D V N P A L A E O K R A E B A V A A V M A S C P 1293  
 1234 L I Q I F F F R L N L I T L L G L H G L V F L P V I L S Y G P D V N P A L A E O K R A E B A V A A V M A S C P 1293  
 DB 1234 L I Q I F F F R L N L I T L L G L H G L V F L P V I L S Y G P D V N P A L A E O K R A E B A V A A V M A S C P 1293

QY 1294 N H P S R V S T A D N I Y V N H S F E G S I K G A G A I S N F L P N N G R O F 1332  
 1294 N H P S R V S T A D N I Y V N H S F E G S I K G A G A I S N F L P N N G R O F 1332  
 DB 1294 N H P S R V S T A D N I Y V N H S F E G S I K G A G A I S N F L P N N G R O F 1332

1321 N H P S R V S T A D N I Y V N H S F E G S I K G A G A I S N F L P N N G R O F 1359

## RESULT 7

US-10-663-208A-44  
 ; Sequence 44, Application US/10663208A  
 ; Publication No. US20040132058A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Altman, Scott W  
 ; APPLICANT: Wang, Luquan  
 ; APPLICANT: Graziano, Michael  
 ; APPLICANT: Murgolo, Nick  
 ; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF  
 ; FILE REFERENCE: JB01603K2 US  
 ; CURRENT APPLICATION NUMBER: US/10/663,208A  
 ; CURRENT FILING DATE: 2003-09-16  
 ; PRIOR APPLICATION NUMBER: 60/397,442  
 ; PRIOR FILING DATE: 2002-07-19  
 ; PRIOR APPLICATION NUMBER: 10/621,758  
 ; PRIOR FILING DATE: 2003-07-17  
 ; PRIOR APPLICATION NUMBER: 10/646,301  
 ; PRIOR FILING DATE: 2003-08-22  
 ; NUMBER OF SEQ ID NOS: 50  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 44  
 ; LENGTH: 1359  
 ; TYPE: prt  
 ; ORGANISM: Homo sapiens  
 ; US-10-663-208A-44

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;  
 Best Local Similarity 97.9%; Pred. No. 0;  
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY 1 M E A G I R G W L M A L L I R L A O S E P Y T T I H O P G Y C A F Y D E C G K N P E L S G S I M T L S N V S C L S N 60  
 1 M A E A G I R G W L M A L L I R L A O S E P Y T T I H O P G Y C A F Y D E C G K N P E L S G S I M T L S N V S C L S N 60  
 DB 1 M A E A G I R G W L M A L L I R L A O S E P Y T T I H O P G Y C A F Y D E C G K N P E L S G S I M T L S N V S C L S N 60

QY 61 T P A R K I T G D H L I L L O K I C P R L Y T G P N T O A C S A K O L V S L E A S I S T K A L L T R C P A C S D N F 120  
 61 T P A R K I T G D H L I L L O K I C P R L Y T G P N T O A C S A K O L V S L E A S I S T K A L L T R C P A C S D N F 120  
 DB 61 T P A R K I T G D H L I L L O K I C P R L Y T G P N T O A C S A K O L V S L E A S I S T K A L L T R C P A C S D N F 120

QY 121 V N L H C H N T S P N O S L F I N T R V A Q L G A G O L P A V V A Y E A F Y Q H S F A Q S Y D S C S R V P V P A 180  
 121 V N L H C H N T S P N O S L F I N T R V A Q L G A G O L P A V V A Y E A F Y Q H S F A Q S Y D S C S R V P V P A 180  
 DB 121 V N L H C H N T S P N O S L F I N T R V A Q L G A G O L P A V V A Y E A F Y Q H S F A Q S Y D S C S R V P V P A 180

QY 181 A T L A V G T M C G Y G S A L C N A Q R W L N F O G D T G N G L A P D I T F H L L E P Q A V G S G I O P L N E G V 240  
 181 A T L A V G T M C G Y G S A L C N A Q R W L N F O G D T G N G L A P D I T F H L L E P Q A V G S G I O P L N E G V 240  
 DB 181 A T L A V G T M C G Y G S A L C N A Q R W L N F O G D T G N G L A P D I T F H L L E P Q A V G S G I O P L N E G V 240

QY 241 A R C N E S Q G D V A T C S C O D C A A S C P A I A R P Q A L D S T F Y L G M P G S V L I I I L C S V F A V V T I 300  
 241 A R C N E S Q G D V A T C S C O D C A A S C P A I A R P Q A L D S T F Y L G M P G S V L I I I L C S V F A V V T I 300  
 DB 241 A R C N E S Q G D V A T C S C O D C A A S C P A I A R P Q A L D S T F Y L G M P G S V L I I I L C S V F A V V T I 300

```

Db      241  ARCNESQGDVATCSQDDCAASCALARPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300
Qy      301  LLYGFRVAPARDKSNMVPKKGTSLSLSDKLSFSTHTLLGQFOGKGTVAWMLTILVLSV 360
Db      301  LLYGFRVAPARDKSNMVPKKGTSLSLSDKLSFSTHTLLGQFOGKGTVAWMLTILVLSV 360
Qy      361  IPVVALAAGVFTTELTPDVELMSAPNSOARSEKAFHDQHPFRTRNOVILTAENRSY 420
Db      361  IPVVALAAGVFTTELTPDVELMSAPNSOARSEKAFHDQHPFRTRNOVILTAENRSY 420
Qy      421  RYDSSLGPKNFGSLDLIDLELELELELELELELELELELELELELELELELELELE 480
Db      421  RYDSSLGPKNFGSLDLIDLELELELELELELELELELELELELELELELELELELE 480
Qy      481  SLVDCINSLLQYFQNNRTLLLTANQTLNGQTSQVDMKHFVLCANAPLTFKDGITAL 540
Db      481  SLVDCINSLLQYFQNNRTLLLTANQTLNGQTSQVDMKHFVLCANAPLTFKDGITAL 540
Qy      541  SCMAVYGAVPFPFLAIGYKGDYSEABALIMTSLNNYPAGDRLAOKLMEBAFLEEM 600
Db      541  SCMAVYGAVPFPFLAIGYKGDYSEABALIMTSLNNYPAGDRLAOKLMEBAFLEEM 600
Qy      601  RAFORBMAGMFQVTFEASRLDEINRTTAEDLPFATSYIVIFLISLALGSYSMSRV 660
Db      601  RAFORBMAGMFQVTFEASRLDEINRTTAEDLPFATSYIVIFLISLALGSYSMSRV 660
Qy      661  MVDKATLGLGVAVVVGAVMAAGFFSYLGRSSLVILQVVPFLVLSVGADNIFILE 720
Db      661  MVDKATLGLGVAVVVGAVMAAGFFSYLGRSSLVILQVVPFLVLSVGADNIFILE 720
Qy      721  YQRLPRRGEPREVHIGRAGRVAPSMULCSLSEALICEFICALTPMPAVRTPALTSGLA 780
Db      721  YQRLPRRGEPREVHIGRAGRVAPSMULCSLSEALICEFICALTPMPAVRTPALTSGLA 780
Qy      781  ILDFLQMSAFVALLISLDSKQEASRLDYCCVCKQELPRPGQSGLLGFFQXAYAFEL 840
Db      781  ILDFLQMSAFVALLISLDSKQEASRLDYCCVCKQELPRPGQSGLLGFFQXAYAFEL 840
Qy      841  LHMVTRGVVLLFLALFGVSLYSMCHISVGLDQELAPKDSYLDLFLMLRYPFVGAPV 900
Db      841  LHMVTRGVVLLFLALFGVSLYSMCHISVGLDQELAPKDSYLDLFLMLRYPFVGAPV 900
Qy      901  YFVMTLGNPSSEAGMNAICSSAGCNMPSFTOKIYATEFPPOSYLAIPASSWVDFIDM 960
Db      901  YFVMTLGNPSSEAGMNAICSSAGCNMPSFTOKIYATEFPPOSYLAIPASSWVDFIDM 960
Qy      961  LTPSSCCRLYISGPNKDFCPESTVNSLNCIKNCSITMGVSVPSEVQFHXYLPWFLNDRP 1020
Db      961  LTPSSCCRLYISGPNKDFCPESTVNSLNCIKNCSITMGVSVPSEVQFHXYLPWFLNDRP 1020
Qy      1021  NIKCPKGLAAYSTSVNLTSDQVLT-----ASRPMAYH 1053
Db      1021  NIKCPKGLAAYSTSVNLTSDQVLT-----ASRPMAYH 1053
Qy      1054  KPLKNSQDYTEALRAARELIANITADLRKVPGTDPAFEPYTTINVPFEOYLITLPEGL 1113
Db      1054  KPLKNSQDYTEALRAARELIANITADLRKVPGTDPAFEPYTTINVPFEOYLITLPEGL 1113
Qy      1114  FMLSLCTVPTFAVSCLLGLDRLSGILNLSITVMLVDVVGFMALMDISYNVSLINLVS 1173
Db      1114  FMLSLCTVPTFAVSCLLGLDRLSGILNLSITVMLVDVVGFMALMDISYNVSLINLVS 1173
Qy      1141  FMLSLCTVPTFAVSCLLGLDRLSGILNLSITVMLVDVVGFMALMDISYNVSLINLVS 1200
Db      1141  FMLSLCTVPTFAVSCLLGLDRLSGILNLSITVMLVDVVGFMALMDISYNVSLINLVS 1200
Qy      1174  ANGMSEVFSHTTRSPALSTKPTMLERAKEATISMGSAVPAGVAMTNLPGIILVGLAKQ 1233
Db      1201  ANGMSEVFSHTTRSPALSTKPTMLERAKEATISMGSAVPAGVAMTNLPGIILVGLAKQ 1260
Qy      1234  LLOIFFERLNTLLITLGLHGLVFLVILSYGPDVNPALALBOKRABEAVAAVWVASC 1293
Db      1261  LLOIFFERLNTLLITLGLHGLVFLVILSYGPDVNPALALBOKRABEAVAAVWVASC 1320
Qy      1294  NHPSRVSTADNITYVNSHFGSITGAGALSNFLPNNGROF 1332

```

```

Db      1321  NHPSRVSTADNITYVNSHFGSITGAGALSNFLPNNGROF 1359

RESULT 8
US-10-646-301A-44
; Sequence 44, Application US/10646301A
; Publication NO. US20040137467A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Margolo, Nick
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: J01603-K1-US
; CURRENT APPLICATION NUMBER: US/10/646,301A
; CURRENT FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 1359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-646-301A-44

Query Match      99.5%; Score 6872.5; DB 4; Length 1359;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

Qy      1  MAEGLGMLIMALLFLAQSERTTHQGYCAFYDECCKNPLSGSLMTLSNVSCLSN 60
Db      1  MAEGLGMLIMALLFLAQSERTTHQGYCAFYDECCKNPLSGSLMTLSNVSCLSN 60
Qy      61  TPARKITGDHILLOKICPRLYTGPNTQACCSAKQVLSLEASTITRALTRCPACSNF 120
Db      61  TPARKITGDHILLOKICPRLYTGPNTQACCSAKQVLSLEASTITRALTRCPACSNF 120
Qy      121  VNLCHNTCSPNQSLFINTVTRVAQLGAGQLPAAVVAEAFYQHSFAEQSYDSCSRVRPAA 180
Db      121  VNLCHNTCSPNQSLFINTVTRVAQLGAGQLPAAVVAEAFYQHSFAEQSYDSCSRVRPAA 180
Qy      121  VNLCHNTCSPNQSLFINTVTRVAQLGAGQLPAAVVAEAFYQHSFAEQSYDSCSRVRPAA 180
Db      121  VNLCHNTCSPNQSLFINTVTRVAQLGAGQLPAAVVAEAFYQHSFAEQSYDSCSRVRPAA 180
Qy      181  ATLAVGTCGYGSGALCNAQRMNLPQDGTNGALPLDITTHLBPQAVSGGIQPLNCGV 240
Db      181  ATLAVGTCGYGSGALCNAQRMNLPQDGTNGALPLDITTHLBPQAVSGGIQPLNCGV 240
Qy      241  ARCNESQGDVATCSQDDCAASCALARPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300
Db      241  ARCNESQGDVATCSQDDCAASCALARPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300
Qy      241  ARCNESQGDVATCSQDDCAASCALARPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300
Db      241  ARCNESQGDVATCSQDDCAASCALARPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300
Qy      301  LLYGFRVAPARDKSNMVPKKGTSLSLSDKLSFSTHTLLGQFOGKGTVAWMLTILVLSV 360
Db      301  LLYGFRVAPARDKSNMVPKKGTSLSLSDKLSFSTHTLLGQFOGKGTVAWMLTILVLSV 360
Qy      361  IPVVALAAGVFTTELTPDVELMSAPNSOARSEKAFHDQHPFRTRNOVILTAENRSY 420
Db      361  IPVVALAAGVFTTELTPDVELMSAPNSOARSEKAFHDQHPFRTRNOVILTAENRSY 420
Qy      421  RYDSSLGPKNFGSLDLIDLELELELELELELELELELELELELELELELELELELE 480
Db      421  RYDSSLGPKNFGSLDLIDLELELELELELELELELELELELELELELELELELELE 480
Qy      481  SLVDCINSLLQYFQNNRTLLLTANQTLNGQTSQVDMKHFVLCANAPLTFKDGITAL 540
Db      481  SLVDCINSLLQYFQNNRTLLLTANQTLNGQTSQVDMKHFVLCANAPLTFKDGITAL 540
Qy      541  SCMAVYGAVPFPFLAIGYKGDYSEABALIMTSLNNYPAGDRLAOKLMEBAFLEEM 600
Db      541  SCMAVYGAVPFPFLAIGYKGDYSEABALIMTSLNNYPAGDRLAOKLMEBAFLEEM 600
Qy      601  RAFORBMAGMFQVTFEASRLDEINRTTAEDLPFATSYIVIFLISLALGSYSMSRV 660

```

```
Db 601 RAFORRMAGMFOVTTFAERSLEDEINRTTAEDLPITATSYIYIFLISLALSYSMSRV 660
Qy 661 MVDSKATIGLGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Db 661 MVDSKATIGLGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Qy 721 YORLPRRPGEPREVIHGRALGRVAPSMILCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Db 721 YORLPRRPGEPREVIHGRALGRVAPSMILCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Qy 781 ILDFLLQMSAFVALSLSDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFPOKAYAPFL 840
Db 781 ILDFLLQMSAFVALSLSDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFPOKAYAPFL 840
Qy 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLDYFLFLNRYFEVAVP 900
Db 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLDYFLFLNRYFEVAVP 900
Qy 901 YFVTTLGYNFSSBAGMNAICSSAGCNSFTOKIQYATEPPEQSYLAIPASSVWDDFIDW 960
Db 901 YFVTTLGYNFSSBAGMNAICSSAGCNSFTOKIQYATEPPEQSYLAIPASSVWDDFIDW 960
Qy 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCCLKMSITWGSVRPSVEQPHKTYLPMFLNDRP 1020
Db 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCCLKMSITWGSVRPSVEQPHKTYLPMFLNDRP 1020
Qy 1021 NIKCPKGLAAYSTSVNLTSDDQVL-----ASRFMAHY 1053
Db 1021 NIKCPKGLAAYSTSVNLTSDDQVL-----ASRFMAHY 1053
Qy 1054 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAEVPYTTNVPFQYITLIPEG 1113
Db 1054 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAEVPYTTNVPFQYITLIPEG 1113
Qy 1114 PMLSLCLVPTFAVSCLLGLDLSGLNLSTWMLVDVGFALMGISTNAASLNLNLS 1173
Db 1114 PMLSLCLVPTFAVSCLLGLDLSGLNLSTWMLVDVGFALMGISTNAASLNLNLS 1173
Qy 1174 AVGMSVEPVSHITRSFAISTKPTWLERAKEATISMSAVFAGVAMTNLPGLVLGLAKQ 1233
Db 1201 AVGMSVEPVSHITRSFAISTKPTWLERAKEATISMSAVFAGVAMTNLPGLVLGLAKQ 1260
Qy 1234 LIQIIEFFRLNLITLGLHGLVFLPVIIISYGPDPVPALALFQKAEBAVAAVMVASC 1293
Db 1261 LIQIIEFFRLNLITLGLHGLVFLPVIIISYGPDPVPALALFQKAEBAVAAVMVASC 1320
Qy 1294 NHPSRVSTADNTVYNSFEFSIKGAGISNLEPNNGROF 1332
Db 1321 NHPSRVSTADNTVYNSFEFSIKGAGISNLEPNNGROF 1359

RESULT 9
US-10-736-769-44
; Sequence 44, Application US/10736769
; Publication No. US20040161838A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: J01603-K3-US
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US/10/736,769
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/546,301
; PRIOR FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: 10/663,208
; PRIOR FILING DATE: 2003-09-16
```

```
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 44
; LENGTH: 1359
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-736-769-44

Query Match          99.5%; Score 6872.5; DB 4; Length 1359;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

Qy 1 MABAGRGWLMALRLRLAOSPEYTHIQPGCAFYDEGKRPETSGSLMTLSNVCISLN 60
Db 1 MABAGRGWLMALRLRLAOSPEYTHIQPGCAFYDEGKRPETSGSLMTLSNVCISLN 60
Qy 61 TPARKITGDHLILQICPRLTYGTGNTQACCSAKQVLSLEASTIKALLTRCPACSDNF 120
Db 61 TPARKITGDHLILQICPRLTYGTGNTQACCSAKQVLSLEASTIKALLTRCPACSDNF 120
Qy 121 VNLHCHNTCSPNQSLFINTVRAQLGAGQLPAVVA YEAFYQHSFABQSYDSCSRVRVPA 180
Db 121 VNLHCHNTCSPNQSLFINTVRAQLGAGQLPAVVA YEAFYQHSFABQSYDSCSRVRVPA 180
Qy 181 ATLAVGTMGVYGSALCNARWLNFOGDTGNGCLAPLDTFHLLEPGQANGSGLIPLNEGV 240
Db 181 ATLAVGTMGVYGSALCNARWLNFOGDTGNGCLAPLDTFHLLEPGQANGSGLIPLNEGV 240
Qy 241 ARCNSQGDVATCGCDCAACCPAIPALDSTFYLGQMGSLVLIITLGSFAVVTI 300
Db 241 ARCNSQGDVATCGCDCAACCPAIPALDSTFYLGQMGSLVLIITLGSFAVVTI 300
Qy 301 LLVGRFVAPARDKSKRVBPKKGTSLSDKLSFSTHTLLGQFFQGWGTWASWPLTILVSV 360
Db 301 LLVGRFVAPARDKSKRVBPKKGTSLSDKLSFSTHTLLGQFFQGWGTWASWPLTILVSV 360
Qy 361 IPVVALAGLVFTELTTPPYELMSA PMSQABSEKA FHQHPGPFRTNTOVILTA PRSSY 420
Db 361 IPVVALAGLVFTELTTPPYELMSA PMSQABSEKA FHQHPGPFRTNTOVILTA PRSSY 420
Qy 421 RYDSSLIGPKNSGILDLLELELOERLHLQVMSPEAQRNLSIODICVAPLPNDT 480
Db 421 RYDSSLIGPKNSGILDLLELELOERLHLQVMSPEAQRNLSIODICVAPLPNDT 480
Qy 481 SLYDCCINSLOYFQNNRTLLLTANQTMGQTSQVDWKDHLFYCANAPLTKDGTALAL 540
Db 481 SLYDCCINSLOYFQNNRTLLLTANQTMGQTSQVDWKDHLFYCANAPLTKDGTALAL 540
Qy 541 SCMAQYGAVPFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQKLMEEAFLERM 600
Db 541 SCMAQYGAVPFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQKLMEEAFLERM 600
Qy 601 RAFORRMAGMFOVTTFAERSLEDEINRTTAEDLPITATSYIYIFLISLALSYSMSRV 660
Db 601 RAFORRMAGMFOVTTFAERSLEDEINRTTAEDLPITATSYIYIFLISLALSYSMSRV 660
Qy 661 MVDSKATIGLGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Db 661 MVDSKATIGLGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Qy 721 YORLPRRPGEPREVIHGRALGRVAPSMILCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Db 721 YORLPRRPGEPREVIHGRALGRVAPSMILCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Qy 781 ILDFLLQMSAFVALSLSDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFPOKAYAPFL 840
Db 781 ILDFLLQMSAFVALSLSDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFPOKAYAPFL 840
Qy 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLDYFLFLNRYFEVAVP 900
Db 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLDYFLFLNRYFEVAVP 900
Qy 901 YFVTTLGYNFSSBAGMNAICSSAGCNSFTOKIQYATEPPEQSYLAIPASSVWDDFIDW 960
```

Db 901 YFTTIGYNFSSAGMNAICSSAGCNPFSTQKIQTATEPPEQSYLAPASSWDDPFIW 960  
Qy 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCJNCMSITMGSVRSVEQFHKYLPWFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCJNCMSITMGSVRSVEQFHKYLPWFLNDRP 1020  
Qy 1021 NIKCPKGLAAVYSTVNLSDGVL-----ASRFMAVH 1053  
Db 1021 NIKCPKGLAAVYSTVNLSDGVLDTVALISPLREYSGTISAHCNLYLSDASRFMAVH 1080  
Qy 1054 KPLKNSODTTEALRAARELANITADLRKPGSDPAEVEPPYTTINVFYQYLTILPEGL 1113  
Db 1081 KPLKNSODTTEALRAARELANITADLRKPGSDPAEVEPPYTTINVFYQYLTILPEGL 1140  
Qy 1114 FMLSCLVPTFAVSCLLIGDLRSGLNLISLIVILVDVGFMAIMDISYNAVSLNLVLS 1173  
Db 1141 FMLSCLVPTFAVSCLLIGDLRSGLNLISLIVILVDVGFMAIMDISYNAVSLNLVLS 1200  
Qy 1174 AVGMSVEPVSHITRSPFAISTKPTWLEBAKEATISMGSAVPAGVAMTNLPGLIVLGLAKQ 1233  
Db 1201 AVGMSVEPVSHITRSPFAISTKPTWLEBAKEATISMGSAVPAGVAMTNLPGLIVLGLAKQ 1260  
Qy 1234 LIOIFFRRLNLITLGLHGLVFLPYILSYGPDVNPALALPQKRAEEVAAMVMAACP 1293  
Db 1261 LIOIFFRRLNLITLGLHGLVFLPYILSYGPDVNPALALPQKRAEEVAAMVMAACP 1320  
Qy 1294 NHPRSVSTADNIVYNSFEGSISGAGAISNPLPNNGROF 1332  
Db 1321 NHPRSVSTADNIVYNSFEGSISGAGAISNPLPNNGROF 1359

RESULT 10  
US-10-450-763-53052  
; Sequence 53052, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 53052  
; LENGTH: 1344  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (901)..(917)  
; OTHER INFORMATION: OESTROGEN RECEPTOR SIGNATURE domain identified by eMATRIX.  
; OTHER INFORMATION: accession number PR00543D, p-value=1.355e-09, raw score of 10.87  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1344)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-53052

Query Match 94.6%; Score 6536; DB 5; Length 1344;  
Best Local Similarity 96.7%; Pred. No. 0;  
Matches 1271; Conservative 7; Mismatches 21; Indels 16; Gaps 4;

Qy 1 MEAGRGMTLMLRLIOSEPYTTHOBYGCAFYDEGSKNPISGSLMTTSSNVCLSN 60  
Db 1 MEAGRGMTLMLRLIOSEPYTTHOBYGCAFYDEGSKNPISGSLMTTSSNVCLSN 60

Qy 61 TPAKRTGDHLILLOKICRPLTYTGPNTOACCSAKOLVSLPASTIKTALLTRCPACSDNF 120  
Db 61 TPAKRTGDHLILLOKICRPLTYTGPNTOACCSAKOLVSLPASTIKTALLTRCPACSDNF 120  
Qy 121 VNIHCHNTCSPNOSLFINTRVAGOLPAVVAEAFQHSFAEOSYDSCSRVPPAA 180  
Db 121 VNIHCHNTCSPNOSLFINTRVAGOLPAVVAEAFQHSFAEOSYDSCSRVPPAA 180  
Qy 181 ATLAATMGCVYSALCNARWLNFOGDTGNGIAPLIDITPHLLPEQAVSGIOPLINEGV 240  
Db 181 ATLAATMGCVYSALCNARWLNFOGDTGNGIAPLIDITPHLLPEQAVSGIOPLINEGV 240  
Qy 241 ARCNSQGDVATCSQDDCAACP--ALAPQALDSTFYLGQMPGSLVLIILCSVAVV 238  
Db 241 ARCNSQGDVATCSQDDCAACP--ALAPQALDSTFYLGQMPGSLVLIILCSVAVV 238  
Qy 299 TILVGFVRVAPADKSMVDPKKGTSISDKLSSTHTLLGQFPGMGTVWASMPLTLYL 358  
Db 299 TILVGFVRVAPADKSMVDPKKGTSISDKLSSTHTLLGQFPGMGTVWASMPLTLYL 358  
Qy 359 SVIPVVALAAGLVFTELTTDPVELMSAPNSQASEKAFDQHPGFRTNOVILTPANRS 418  
Db 359 SVIPVVALAAGLVFTELTTDPVELMSAPNSQASEKAFDQHPGFRTNOVILTPANRS 418  
Qy 419 SYRYDSSLGPKNFGSILDLILBELLEOERLRHQQVSPBAQRNISLDDICYAPLND 478  
Db 419 SYRYDSSLGPKNFGSILDLILBELLEOERLRHQQVSPBAQRNISLDDICYAPLND 478  
Qy 479 NTSLYOCCNSLIQYQONNRITLILANOTIMQTSQVMDKHFLCANAPLTFKQGTAL 538  
Db 479 NTSLYOCCNSLIQYQONNRITLILANOTIMQTSQVMDKHFLCANAPLTFKQGTAL 538  
Qy 539 ALSCMADYGAPEPFLAIGYKGYKDYSEAEALIMTESLNNYPADPRLAQAKLMEAEFL 538  
Db 539 ALSCMADYGAPEPFLAIGYKGYKDYSEAEALIMTESLNNYPADPRLAQAKLMEAEFL 538  
Qy 599 EKRAFORMAQMFQVTFABERSLDEINRTTAEPLIPATSYIVILYISLALGSYSWS 658  
Db 599 EKRAFORMAQMFQVTFABERSLDEINRTTAEPLIPATSYIVILYISLALGSYSWS 658  
Qy 659 RNVVDSKATLGLGAVAVVGAWMAAGFSSYIGIRSSVILQVNPVLNVSGADNIFIV 718  
Db 659 RNVVDSKATLGLGAVAVVGAWMAAGFSSYIGIRSSVILQVNPVLNVSGADNIFIV 718  
Qy 719 LEYQRLPRRPGEBREHIGRALGRVAPSMULCSLSAICFFLGALTPMVRKFTALTSGL 778  
Db 719 LEYQRLPRRPGEBREHIGRALGRVAPSMULCSLSAICFFLGALTPMVRKFTALTSGL 778  
Qy 779 AVILDFLQMSAFVALLSLDSKROEASRLDVCCVPRQELPPRGQEGILLGFFQKAYAP 838  
Db 779 AVILDFLQMSAFVALLSLDSKROEASRLDVCCVPRQELPPRGQEGILLGFFQKAYAP 838  
Qy 839 FLHMTTRGVV---LLFLALFGVSLYGMCHISVGLDDELALPKOSYLLDYLFPLNRYF 894  
Db 839 FLHMTTRGVVPSQLLFLALFGVSLYGMCHISVGLDDELALPKOSYLLDYLFPLNRYF 898  
Qy 895 EVGAPYFTTIGYNFSSAGMNAICSSAGCNPFSTQKIQTATEPPEQSYLAPASSWV 954  
Db 895 EVGAPYFTTIGYNFSSAGMNAICSSAGCNPFSTQKIQTATEPPEQSYLAPASSWV 958  
Qy 955 DPEIDWLTPSSCCRLYISGPNKDKFCPSTVNSLNCJNCMSITMGSVRSVEQFHKYLPW 1014  
Db 955 DPEIDWLTPSSCCRLYISGPNKDKFCPSTVNSLNCJNCMSITMGSVRSVEQFHKYLPW 1018  
Qy 1015 FLNDRPNKCPKGLAAVYSTVNLSDGVLASRFMAVYHKPLKNSODYTEALRAARELAA 1074  
Db 1015 FLNDRPNKCPKGLAAVYSTVNLSDGVLASRFMAVYHKPLKNSODYTEALRAARELAA 1078  
Qy 1075 NITADLRKVPGSDPAEVEPPYTTINVFYQYLTILPEGLFMLSCLVPTFAVSCLLIGLD 1134  
Db 1075 NITADLRKVPGSDPAEVEPPYTTINVFYQYLTILPEGLFMLSCLVPTFAVSCLLIGLD 1138  
Qy 1135 LRSGLNLISLIVILVDVGFMAIMDISYNAVSLNLVSAVGMSVEPVSHITRSPFAISTK 1194

```

Db      1139  LRSGLNLSIWMILVDYGFALNGISYNAVSLINLSAVGMSVEFVSHIRSFATSRK 1198
Oy      1195  PFWLBRRAKATISMGSAVAGVAMTNLPGLILVLGLAKAOLIOIFFRLNLTITLGLLHG 1254
Db      1199  PFWLBRRAKATISMGSAVAGVAMTNLPGLILVLGLAKAOLIOIFFRLNLTITLGLLHG 1258
Oy      1255  LVFLPVILSYGPDVNPALALBOKRAEBA-----VAAMVASCPNHRPVRST 1301
Db      1259  LVFLPVILSYGPDVNPALALBOKRAEBAEGGSGHGPLAQITPBPSPQLTTSMTST 1313

RESULT 11
US-10-621-758A-2
; Sequence 2, Application US/10621758A
; Publication No. US20040093629A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-K-US
; CURRENT APPLICATION NUMBER: US/10/621,758A
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PRF
; ORGANISM: Rattus sp.
US-10-621-758A-2

Query Match      78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

Oy      1  MEEAGLRGWLMAALLRLAOSPEYTTIHQGYCAFYDECGKNPELGSIMTISNVASCLSN 60
Db      1  MAAAMU-GWLLMNLNLSSAAOGEIYTPKHEAGVCTFEECGKNPELGSGLTSLNSVCLSN 59
Oy      61  TPAKRTGTGHTLLQKICPRLYTGPNTQ-ACCSAKOLVSLBASLTITKALLTRCPACSDN 119
Db      60  TPABHTYGHLLAQKICPRLYNGPNTTPACCTKQLSLSESMSITKALLTRCPACSDN 119
Oy      120  FVNLHCHNTCSPPQSLFIVNTRVAOLGAGOLPAVVAVEAFYQHSFPAEGSYDCSRVAPA 179
Db      120  FVSLHCHNTCSPPQSLFIVNTRVERGABEPRAVVAVEAFYQHSFPAEKAYESCSQVRIPA 179
Oy      180  AATLAVGTGCVYVGSALCNARWLNFGQGTNGLAFLADITFHLBEGQAVSGIOPLNKG 239
Db      180  AASLAVGSMGVYVGSALCNARWLNFGQGTNGLAFLADITFHLBEGQALPDGIOPLNKG 239
Oy      240  VARNESQGDVATCSQDCAASCAPAIARPOALDSTFYLGOMPGSLVLIILCSVAVVT 299
Db      240  IAPCNESQGDVATCSQDCAASCAPAIARPOALDSTFYLGOMPGSLVLIILCSVAVVT 299
Oy      300  ILLVGRVAPADKSMVDPKGTSLSDKLSFSTHTLLQFQFGWGTWVASNPITLVLS 359
Db      300  AVLVRARVVSNNRNKNABEGPOEAPKLRPHKHLSPHTILRFQNMGTFRVAVSPLTYLALS 359
Oy      360  VLPVVALAGLVFTELTTPDVELWASAPNSQARSEKAFHNDHGFPRFTQVITLAPNRS 419
Db      360  FIVVVALAGLVFTELTTPDVELWASAPNSQARSEKAFHNDHGFPRFTQVITLAPNRS 419
Oy      420  VYVDSLLGPKVFSGLIDLLLELLELQERLHLQVMSPEAQRNLSLODICVAPLPNP 479
Db      420  VYVDSLLGPKVFSGLISDLLELLELQERLHLQVMSPEAQRNLSLODICVAPLPNP 479
Oy      480  TSLVDCINSLLOYFONNRTLLILTNQITMGTSQVDMKHFLYCANAPLTFKQGTALA 539
Db      480  TSLVDCINSLLOYFONNRTLLILTNQITMGTSQVDMKHFLYCANAPLTFKQGTALA 539

```

```

Db      480  TSLVDCINSLLOYFONNRTLLILTNQITMGTSQVDMKHFLYCANAPLTFKQGTALA 539
Oy      540  LSCMADYGAPEPFLAIGYKGDYSEAEALIMFESLNNYPAGDRBLAQAKIMEAPFEE 599
Db      540  LSCMADYGAPEPFLAIGYKGDYSEAEALIMFESLNNYPAGDRBLAQAKIMEAPFEE 599
Oy      600  MRAFORNAGMPOVTFTEKERSLEDEINNTAEDLPFATSYIVITFLYISLALGSSMSR 659
Db      600  MESFORNAGMPOVTFTEKERSLEDEINNTAEDLPFATSYIVITFLYISLALGSSMSR 659
Oy      660  VVWDSKATLIGGVAVVIGAAVMAAMGFPFSYIGIRSLYILOVPPVLYSVGADNIFIVL 719
Db      660  VVWDSKATLIGGVAVVIGAAVMAAMGFPFSYIGIRSLYILOVPPVLYSVGADNIFIVL 719
Oy      720  EYORLPRRBGEVREHIGRALGRVAPSMULCSLSBAICFPGLATPMPAVRTFALTSGIA 779
Db      720  EYORLPRRBGEVREHIGRALGRVAPSMULCSLSBAICFPGLATPMPAVRTFALTSGIA 779
Oy      780  VILDLQMSAFVALLSLDSKQKASRLDYCCCVKPOELPPRGQESGILLGFQKAYAPF 839
Db      780  IILDLQMTAFVALLSLDSKQKASRPDYCCCVKPOELPPRGQESGILLGFQKAYAPF 839
Oy      840  LLHWITRGVLLFLALPGVSLYSWCHT SVGLDQELAPKDSYLDYFLFLNRFEVGPAP 899
Db      840  LLHWITRGVLLFLALPGVSLYSWCHT SVGLDQELAPKDSYLDYFLFLNRFEVGPAP 899
Oy      900  VYFVTTLLGYNFSSEAGMNAICSSACCNPFSTOKIQVATEPEEGSYLAIIPASSWVDFID 959
Db      900  VYFVTTLLGYNFSSEAGMNAICSSACCNPFSTOKIQVATEPEEGSYLAIIPASSWVDFID 959
Oy      960  WITP-SSCCRLYTSGPNKDKFCPTSVNSLNCILKAKMSITMGSVRSVQOFKXFLPWFIND 1018
Db      960  WITP-SSCCRLYTSGPNKDKFCPTSVNSLNCILKAKMSITMGSVRSVQOFKXFLPWFIND 1018
Oy      1019  RPNIKCPKGLAAAYSTVNLSDGOVLASRFMAHYKPKJNSODDYTEALRAARELANITTA 1078
Db      1020  RPNIKCPKGLAAAYSTVNLSDGOVLASRFMAHYKPKJNSODDYTEALRAARELANITTA 1078
Oy      1079  DLKVPGTDPAPFEPYTTITVNFYBOYLTILPEBGLFMLSCLVPTFAVSCLLGLDLBSG 1138
Db      1080  DLKVPGTDPAPFEPYTTITVNFYBOYLTILPEBGLFMLSCLVPTFAVSCLLGLDLBSG 1138
Oy      1139  LNLISLIMILVDITGFMALNDISYNAVSLINLSAVGMSVEFVSHIRSFATSRKPTWL 1198
Db      1140  LNLISLIMILVDITGFMALNDISYNAVSLINLSAVGMSVEFVSHIRSFATSRKPTWL 1198
Oy      1199  ERAKATISMGSAVAGVAMTNLPGLILVLGLAKAOLIOIFFRLNLTITLGLLGLVFL 1258
Db      1200  ERAKATISMGSAVAGVAMTNLPGLILVLGLAKAOLIOIFFRLNLTITLGLLGLVFL 1258
Oy      1259  PVILSYGPDVNPALALBOKRAEBAVMAVMAVMAVMAVMAVMAVMAVMAVMAVMAVMAV 1317
Db      1260  PVILSYGPDVNPALALBOKRAEBAVMAVMAVMAVMAVMAVMAVMAVMAVMAVMAVMAV 1316
Oy      1318  AGAISNPLPNNGROF 1332
Db      1317  ANAARSPLPKSDOKF 1331

RESULT 12
US-10-663-208A-2
; Sequence 2, Application US/10663208A
; Publication No. US20040132058A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603K2 US
; CURRENT APPLICATION NUMBER: US/10/663,208A
; CURRENT FILING DATE: 2003-09-16
; PRIOR APPLICATION NUMBER: 60/397,442

```



```

; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/646,301
; PRIOR FILING DATE: 2003-08-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PR
; ORGANISM: Rattus sp.
US-10-663-208A-2

```

```

Query Match      78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

```

```

QY 1 MAEAGRGMLLMALLRLAOSBPYTHHQGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 60
DB 1 MAEAGRGMLLMALLRLAOSBPYTHHQGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 59
QY 61 TPAARKITGDHLLLOKICPRLYNGPNTQ-ACCSAKOLVSLBASISITKALLTRCPACSDN 119
DB 60 TPAARKITGDHLLLOKICPRLYNGPNTQ-ACCSAKOLVSLBASISITKALLTRCPACSDN 119
QY 120 FVNLHCNHTCSPNOSLFINTRVAQLGAGOLPAVAVAEAFYOHSPAEQSDSCSRVAVPA 179
DB 120 FVNLHCNHTCSPNOSLFINTRVAQLGAGOLPAVAVAEAFYOHSPAEQSDSCSRVAVPA 179
QY 120 FVNLHCNHTCSPNOSLFINTRVAQLGAGOLPAVAVAEAFYOHSPAEQSDSCSRVAVPA 179
DB 120 FVNLHCNHTCSPNOSLFINTRVAQLGAGOLPAVAVAEAFYOHSPAEQSDSCSRVAVPA 179
QY 180 AATLAVGTGCVYGSALCNQRMWLNFGDGTNGIAPLDITFHLLPEQAVSGGIQPLNEG 239
DB 180 AATLAVGTGCVYGSALCNQRMWLNFGDGTNGIAPLDITFHLLPEQAVSGGIQPLNEG 239
QY 180 AATLAVGTGCVYGSALCNQRMWLNFGDGTNGIAPLDITFHLLPEQAVSGGIQPLNEG 239
DB 180 AATLAVGTGCVYGSALCNQRMWLNFGDGTNGIAPLDITFHLLPEQAVSGGIQPLNEG 239
QY 240 VACNESQGDVATCSCODCAAPARPAQALDSTFYLGOMPSVLIIITLCSVEAVVT 299
DB 240 VACNESQGDVATCSCODCAAPARPAQALDSTFYLGOMPSVLIIITLCSVEAVVT 299
QY 240 VACNESQGDVATCSCODCAAPARPAQALDSTFYLGOMPSVLIIITLCSVEAVVT 299
DB 240 VACNESQGDVATCSCODCAAPARPAQALDSTFYLGOMPSVLIIITLCSVEAVVT 299
QY 300 ILVGFPRVAPARDKSKVNDPKGTSLSDKLSFTHTLLGQFQMGMTWVASMPLTIIYLS 359
DB 300 ILVGFPRVAPARDKSKVNDPKGTSLSDKLSFTHTLLGQFQMGMTWVASMPLTIIYLS 359
QY 300 AVLVRLVRVNRNKNKAEQOEAPKLPKHKLSPTHTLIGRFQWGRVASMPLTIIYLS 359
DB 300 AVLVRLVRVNRNKNKAEQOEAPKLPKHKLSPTHTLIGRFQWGRVASMPLTIIYLS 359
QY 360 VIVPVALAAGLVTELTTPDPELMSAPNSQARSEKAFHDOHFGPFRTNOVILTAPESS 419
DB 360 VIVPVALAAGLVTELTTPDPELMSAPNSQARSEKAFHDOHFGPFRTNOVILTAPESS 419
QY 360 VIVPVALAAGLVTELTTPDPELMSAPNSQARSEKAFHDOHFGPFRTNOVILTAPESS 419
DB 360 VIVPVALAAGLVTELTTPDPELMSAPNSQARSEKAFHDOHFGPFRTNOVILTAPESS 419
QY 420 YRVDLSLLGPRNPSGIIIDDLLELLELLOERLRLQVMSPEACNITSLQDI CYAPLMPDN 479
DB 420 YRVDLSLLGPRNPSGIIIDDLLELLELLOERLRLQVMSPEACNITSLQDI CYAPLMPDN 479
QY 420 YRVDLSLLGPRNPSGIIIDDLLELLELLOERLRLQVMSPEACNITSLQDI CYAPLMPDN 479
DB 420 YRVDLSLLGPRNPSGIIIDDLLELLELLOERLRLQVMSPEACNITSLQDI CYAPLMPDN 479
QY 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOYDMKDFLYCANAPLTFDGTALA 539
DB 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOYDMKDFLYCANAPLTFDGTALA 539
QY 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOYDMKDFLYCANAPLTFDGTALA 539
DB 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOYDMKDFLYCANAPLTFDGTALA 539
QY 540 LSCMADGAVFPFLAIGYKKGKDYSEAEALIMTFSLNTNYPADGPRLAQAKLMEAEFLER 599
DB 540 LSCMADGAVFPFLAIGYKKGKDYSEAEALIMTFSLNTNYPADGPRLAQAKLMEAEFLER 599
QY 540 LSCMADGAVFPFLAIGYKKGKDYSEAEALIMTFSLNTNYPADGPRLAQAKLMEAEFLER 599
DB 540 LSCMADGAVFPFLAIGYKKGKDYSEAEALIMTFSLNTNYPADGPRLAQAKLMEAEFLER 599
QY 540 LSCMADGAVFPFLAIGYKKGKDYSEAEALIMTFSLNTNYPADGPRLAQAKLMEAEFLER 599
DB 540 LSCMADGAVFPFLAIGYKKGKDYSEAEALIMTFSLNTNYPADGPRLAQAKLMEAEFLER 599
QY 600 MRAFORMAGMFOYTFPERSLEDEINRTTAEDELPATSTIYVFLYISLALGYSKSR 659
DB 600 MRAFORMAGMFOYTFPERSLEDEINRTTAEDELPATSTIYVFLYISLALGYSKSR 659
QY 600 MRAFORMAGMFOYTFPERSLEDEINRTTAEDELPATSTIYVFLYISLALGYSKSR 659
DB 600 MRAFORMAGMFOYTFPERSLEDEINRTTAEDELPATSTIYVFLYISLALGYSKSR 659
QY 600 MRAFORMAGMFOYTFPERSLEDEINRTTAEDELPATSTIYVFLYISLALGYSKSR 659
DB 600 MRAFORMAGMFOYTFPERSLEDEINRTTAEDELPATSTIYVFLYISLALGYSKSR 659
QY 660 VMTDSKATLIGGAVVILGAVMAMGFPSTYIGIRSSLYIIQVAVFVLVSAGADNIFIVL 719
DB 660 VMTDSKATLIGGAVVILGAVMAMGFPSTYIGIRSSLYIIQVAVFVLVSAGADNIFIVL 719
QY 660 VMTDSKATLIGGAVVILGAVMAMGFPSTYIGIRSSLYIIQVAVFVLVSAGADNIFIVL 719
DB 660 VMTDSKATLIGGAVVILGAVMAMGFPSTYIGIRSSLYIIQVAVFVLVSAGADNIFIVL 719
QY 720 EYORLPERRPGEPREVHIGRALGVAPSMULLCSLESEALCFPLGALTTPMVAVTFTALSGLA 779
DB 720 EYORLPERRPGEPREVHIGRALGVAPSMULLCSLESEALCFPLGALTTPMVAVTFTALSGLA 779
QY 720 EYORLPERRPGEPREVHIGRALGVAPSMULLCSLESEALCFPLGALTTPMVAVTFTALSGLA 779
DB 720 EYORLPERRPGEPREVHIGRALGVAPSMULLCSLESEALCFPLGALTTPMVAVTFTALSGLA 779
QY 780 VIIDFLLQMSAFVALLSLDSKROASRLDYCCYKQPELPPGGGEGILLGFPKAYAPF 839
DB 780 VIIDFLLQMSAFVALLSLDSKROASRLDYCCYKQPELPPGGGEGILLGFPKAYAPF 839
QY 780 VIIDFLLQMSAFVALLSLDSKROASRLDYCCYKQPELPPGGGEGILLGFPKAYAPF 839
DB 780 VIIDFLLQMSAFVALLSLDSKROASRLDYCCYKQPELPPGGGEGILLGFPKAYAPF 839

```

```

QY 840 LHMTRGVALLFLALFGVSLYSKCHISVGLDOELALPDOSYLLDYFLUNRYFVGAP 899
DB 840 LHMTRGVALLFLALFGVSLYSKCHISVGLDOELALPDOSYLLDYFLUNRYFVGAP 899
QY 900 VYFPTTLGVNPSSEAGNNAICSSAGCNPSFTOKIOYATFEPBESYALAPASSVNDPFD 959
DB 900 VYFPTTLGVNPSSEAGNNAICSSAGCNPSFTOKIOYATFEPBESYALAPASSVNDPFD 959
QY 900 VYFPTTLGVNPSSEAGNNAICSSAGCNPSFTOKIOYATFEPBESYALAPASSVNDPFD 959
DB 900 VYFPTTLGVNPSSEAGNNAICSSAGCNPSFTOKIOYATFEPBESYALAPASSVNDPFD 959
QY 960 WLP-SSCCRLYISGPNKDKFCSTVNSLNLCKNCSITWGSVAPSVQEPHXYLPWFLND 1018
DB 960 WLP-SSCCRLYISGPNKDKFCSTVNSLNLCKNCSITWGSVAPSVQEPHXYLPWFLND 1018
QY 960 WLP-SSCCRLYISGPNKDKFCSTVNSLNLCKNCSITWGSVAPSVQEPHXYLPWFLND 1018
DB 960 WLP-SSCCRLYISGPNKDKFCSTVNSLNLCKNCSITWGSVAPSVQEPHXYLPWFLND 1018
QY 1019 PENIKCPKGLAAVSTSVNLTSQOVLASFPMAVHKPKNSODYTELRLPARELANITA 1078
DB 1019 PENIKCPKGLAAVSTSVNLTSQOVLASFPMAVHKPKNSODYTELRLPARELANITA 1078
QY 1020 PPMRCPKGLAAVSTSVNLTSQOVLASFPMAVHKPKNSODYTELRLPARELANITA 1079
DB 1020 PPMRCPKGLAAVSTSVNLTSQOVLASFPMAVHKPKNSODYTELRLPARELANITA 1079
QY 1079 DLRKVPCTDPAFEPPTTNVFEQVLTLPBGLPMLSLCLVPTFVSCILLGLDLSRG 1138
DB 1080 DLRKVPCTDPAFEPPTTNVFEQVLTLPBGLPMLSLCLVPTFVSCILLGLDLSRG 1139
QY 1139 LNLSTIWMILVDTVGFMALMDSYNVSLINLVSAVGSVBFVSHITRSPALSTKPTWL 1198
DB 1140 LNLSTIWMILVDTVGFMALMDSYNVSLINLVSAVGSVBFVSHITRSPALSTKPTWL 1199
QY 1199 ERAKATISGSAVAVAGVANTLPGIIVGLAKAQLIQIPFRLNLLITLGLHGVFL 1258
DB 1200 ERAKATISGSAVAVAGVANTLPGIIVGLAKAQLIQIPFRLNLLITLGLHGVFL 1259
QY 1259 PVILSYGPRVNPALMEQKRAEVAANVASCPRNPSVSTADNYYNHSFEGSI-KG 1317
DB 1260 PVILSYGPRVNPALMEQKRAEVAANVASCPRNPSVSTADNYYNHSFEGSI-KG 1316
QY 1318 AGAISNPLPNNNGROF 1332
DB 1317 ANAARSSLPEKSDQKF 1331

```

## RESULT 13

```

US-10-646-301A-2
; Sequence 2, Application US/10646301A
; Publication No. US20040137467A1
; GENERAL INFORMATION:
; APPLICANT: Altmann, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: J01603-K1-US
; CURRENT APPLICATION NUMBER: US/10/646,301A
; PRIOR FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PR
; ORGANISM: Rattus sp.
US-10-646-301A-2

```

```

Query Match      78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

```

```

QY 1 MAEAGRGMLLMALLRLAOSBPYTHHQGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 60
DB 1 MAEAGRGMLLMALLRLAOSBPYTHHQGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 59
QY 61 TPAARKITGDHLLLOKICPRLYNGPNTQ-ACCSAKOLVSLBASISITKALLTRCPACSDN 119
DB 60 TPAARKITGDHLLLOKICPRLYNGPNTQ-ACCSAKOLVSLBASISITKALLTRCPACSDN 119

```



QY	120	FWNHCHNTCSNOSLFTINVTBAQOLPAVVAYEAFYOHSPAEQSDCSRVRVA	179
Db	120	FVSHCHNTCSPDQSLFINVTBVEGAEPRAVYAEAFORSPEKAYESQVRIRP	179
QY	180	AATLAVGTMCGVYGSALCNAORMLNFOGDTGNGLAELDTTFHLEBGAVSGSIQPLNGB	239
Db	180	AASLAVGSMCGVYGSALCNAORMLNFOGDTGNGLAELDTTFHLEBGOALPDIOPLNGB	239
QY	240	VARCHESQDDVAATGSCODCAASCRAIPALDSTFYIGOMGSLVLTILCSVAVVT	299
Db	240	IAPCHESQDDVAATGSCODCAASCRAIPPEALRPSFYGRMGWLTALITTFVAVLVS	299
QY	300	ILLVFRVAPADSKMVDPKKGTSLSDKLSFSTHTLLAQCFQCMGTWYASWPLTIVLS	359
Db	300	AVLVAIRVYNSNKKAKABEGQEARPLPHHKKXLSPHITLDRFQNMKTRVJASWPLTIVLVS	359
QY	360	VIPVVALAAGLVFTELITDPVELMSAPNSOANSERKAFHDOHGFPRTRNQVITADNRSS	419
Db	360	FIVVVALAAGLVFTELITDPVELMSAPKSQARKEKSFHDEHFGPFRTRNOIFVTAANRSS	419
QY	420	YRYSBLLGGPKNPSGILDDLLELELEBERLPHLOVMSPEAORNI SLODICAPLNPN	479
Db	420	YKIDBLLGSKNPSGILSDFLBELLEBERLHLOVMSPEAKRNI SLODICAPLNPN	479
QY	480	TSLYDCCINSLLQYQONNRTELLLTANQTMGOTSQVMDKHELYCANAPLTFKDGSTALA	539
Db	480	TSLSDDCVNSLLQYQONNRTELLMTANQTMGOTSJVDWKDHFLYCANAPLTFKDGSTILA	539
QY	540	LSGMDADYGA PVPFPPLAIGGYKKGKDYSEARALIMTESLNNYPADPRLAOKIMEEAFLE	599
Db	540	LSGMDADYGA PVPFPPLAVGGYQGTDSSEALITTFSLNNYPADPRLAOKIMEEAFLE	599
QY	600	MBAFORBAMGMQVFTFARSLEDEINRTAEDLPFATASYIYIFUYISALGSYASWSR	659
Db	600	MESFORNTSDKQVAFSAERSLEDEINRTTIOPLPFANSYIIVFLYISALGSYASRCSR	659
QY	660	VWVDSKATITGLCGVAVVLGAVMAANGFSYILGRSSLVLTQVVPFLVLSYGADNITFPVL	719
Db	660	VAVESKATITGLCGVAVVLGAVLAAMGFYSYILGPSSVLIVQVVPFLVAVGADNITFPVL	719
QY	720	EYQRLPRRGEPREVHIGALGRVAPSMILCSISEAICFFIALATMPAVRTALNSGLA	779
Db	720	EYQRLPRRGEORERAHIGRTLGSVAPSMILCSISEAICFFIALTIMPAVRTALNSGLA	779
QY	780	VILDEFLQMSAFVALLSDSKQOASRLDVCCCVKQOELPPPGQGBGLLGFPOKAYAP	839
Db	780	IILDEFLQMTAFVALLSDSKQOASRPVULCCFSRTRLPPEKXEGKLLRFPKRTIYAP	839
QY	840	LLHMTITRGVULLFLALRGVSLYSMKHISVGDQOELAKRDSVLTLYFLPLNYPFVGA P	899
Db	840	LLHRTIRPVVULLFLFLFANLYLKMKNINVGDOELAPDOSTLIDYFLPLNYYLEVG P	899
QY	900	VYFVVTLLGYNFSSBAGMNAICSSAGCNFNSFTQKIOYATEFEPQOSYLAI PASSWVDFID	959
Db	900	VYFVVTSGNFSSBAGMATCSAGCKSPSLTKIYASBFPQOSYVALIASSWVDFID	959
QY	960	WITPSSCCRLYISGPNKDKFCPSYVNSLNLCKNCSITMGSYRBSVEBQHKLLPMPFLND	101
Db	960	WLTPESSCCRLYIRGPHKDEFCPSYDTSRCLCNKCMNRITGLVPRAPEBPHKLLPMPFLND	101
QY	1019	RPNICPCKGGLAAYSTVNLTSQGOVLASRFMAVYHPLKNSQOYTEALRAARELANITA	107
Db	1020	PNNICPCKGGLAAYRISVNLSSDGOVIAQOFMAVYHPLKNSQDFTALRAARSLANITA	107
QY	1079	DLRKVPGTDPAEVPEYTTITNVFEQYLTLPBGLFMLSCLVPTFAVSCLLIGDLRSG	113
Db	1080	DLRKVPGTDPNEVEPEYTTISNFYQOYLTVLPBGITFLALCPVPTFVVCYLLIGLDMCSG	113
QY	1139	ILNLISLWIIIVDYJGFMALMDSVNAVSLINVSANVGSVEFVSHITTSFALSTPTWL	119
Db	1140	ILNLISLWIIIVDTJGLMAVWGISVNAVSLINLVTVAGSVSEFVSHITTSFAVSTPTRL	119

```

QY 1199 ERAEAEITISGSAVEAFVAMTNTJPGILVJLGLAKQQLQIFEFRLNTLITLLGLHGVFL 1258
Db 1200 ERADDAVFNQSAVEAFVAMTNTJPGILVJLGLAKQQLQIFEFRLNTLITLLGLHGVFL 1259
QY 1259 PVILSYVGPDPNPALALEOKRAEEVAVAWVASCPNHPSRVSTADNTIYNHSPFGSI-KG 1317
Db 1260 PVVLSTYIGPDVQNLVQEBKLASA-AVADEPSCPOYRPSPADADAN--VNYGFAPELHAG 1316
QY 1318 AGAISNFLPNNGROF 1332
Db 1317 ANAARSSLPKSDOKF 1331

RESULT 14
US-10-736-769-2
; Sequence 2, Application US/10736769
; Publication No. US20040161838A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: J01603-K3-US
; CURRENT APPLICATION NUMBER: US/10/736,769
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/646,301
; PRIOR FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: 10/663,208
; PRIOR FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-736-769-2

Query Match 78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAEAGLGMILLMALLLLAQSEPYTTHOGYCAFYDECGKNPGLSGSLMTLSNVSCISN 60
Db 1 MAAMWL-GMLMALLLLSNAQGEYTPPHKAGVCYTFIECGKNPGLSGSLTSLNVSCISN 59
QY 61 TPARKITGDHLLLOKICPRLTYGPNTO-ACCSAKOLVSLSEASISITKALLTRCPACSDN 119
Db 60 TPAAHYVTEGHEHALLQRCPLRYNGPNNTTPACCSITKOLLSLESSMSITKALLTRCPACSDN 119
QY 120 FVNIHCHNTGSPNOSLFINTTRVAQLGAGQLPAVVAVEAFYQHSFAQOSTDSCARVPA 179
Db 120 FVSLHCHNTGSPDQSLFINTRVVERAGGEPAVVAEAFYQHSFAKAVESCSQVRIPA 179
QY 180 AATLAVGTMGVSSALCNQORWLNPGDGTNGIAPLIDITPHLLEPQOAGSGIOPLNEG 239
Db 180 AASLAVSMGCVSSALCNQORWLNPGDGTNGIAPLIDITPHLLEPQOAGSGIOPLNGK 239
QY 240 VARCNESQGDVAATCSCODCAASCPAIARPOLDSTFYLQGMPSGLVLIILCSVFAVNT 299
Db 240 IAPCNESQGDVAATCSCODCAASCPIVPPBALRPSFTYMGMPGMLLIILIFTVFVLSS 299
QY 300 ILLVGFVAPARDXSRWDPKKGTSLSDKLSFTHTLLAQFGFGWGTWVASWPLTIVLS 359
Db 300 AVLVRLRVSNRNKKAEGPOEAPKLPHKHKLSPHTILIGRFQWNGTRVASWPLTIVALS 359
QY 360 VIVPVALLAAGLVTELTTPPELMSARNQARSKAHHQDFGFFFTNOVITLAPRNS 419
Db 360 FIVVIALAAGLTTELTTPPELMSARQSQAARKKSPHDEHGFGRFTNOITFVARNS 419

```

420 YAYDSLLGPKNPSGILDLLELELELOERLRLHLOVMSPEAORNI SLODI CYAPLNPON 479  
 420 YKXDSLLGKNSGILSLDFLELELELOERLRLHLOVMSPEAORNI SLODI CYAPLNPY 479  
 480 TSLYCCISLLOYPONNRTLLLTANQTLMGOTSOVDKDHLYCANMPLFKDSTALA 539  
 480 TSLSDCVNSLLOYPONNRTLLLTANQTLMGOTSLVDMKDHLYCANMPLFKDSTALA 539  
 540 LSCMDYGAPEPFLAIGGYKDYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLER 599  
 540 LSCMDYGAPEPFLAIGGYKDYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLER 599  
 600 MEAFORHAGMFOVTFPAERSLDEINRTTAEDLPITFATSYIVIFLYISIALGSSYSWSR 659  
 600 MESFOGNTSDKPOVARSARSLEDEINRTTIQDLPPFAVSIIIVFLYISIALGSSYSRCSR 659  
 660 VAWDSKATIGLGAVAVVLAANMAAMGFYLGIRSSLVLIQVPPFLVLSGADNITIFVL 719  
 660 VAVESKATIGLGAVIIVLGAVALAMGFYSYLGVPSSLVIIQVPPFLVLAAGADNIRIFVL 719  
 720 EYORLRRRGEPEPEVHIGRALGRVAPSMILCSLSBAICFPLGALTPMPAVRTALSGLA 779  
 720 EYORLRRRGEPEPEVHIGRALGRVAPSMILCSLSBAICFPLGALTPMPAVRTALSGLA 779  
 780 VILDFLLQMSAFVALSLDSKROEASRLDYCCVKEQOELPPPCQGBGLLGFPOKAYAPF 839  
 780 IILDFLLQMTAFVALSLDSKROEASRPVLCFSTRKLPPEPKKGGLLRRFRKXIYAPF 839  
 840 LHMWITRGVYLLFLALFGVSLYSKCHISVGLDOELAPDOSTYLDLFLANRYEYVGA 899  
 840 LHMWITRGVYLLFLALFGVSLYSKCHISVGLDOELAPDOSTYLDLFLANRYEYVGA 899  
 900 VYFVTTLYGNFSEAGMNAICSSAGCNPSFTQIOYATPEFPROSYLAIPASSWVDFID 959  
 900 VYFVTTLYGNFSEAGMNAICSSAGCNPSFTQIOYATPEFPROSYLAIPASSWVDFID 959  
 960 WLPTR-SSCRLYISGPNKDFCPSYVNSLNCLEKNCMSITWGSVAPSVBOFHXYLPWFLND 1018  
 960 WLPTR-SSCRLYISGPNKDFCPSYVNSLNCLEKNCMSITWGSVAPSVBOFHXYLPWFLND 1018  
 1019 RPNKCPKGGGLAAVSTSVNLTSPGOVLASRFMAVHKLKNSODYTEALRAARLANITA 1078  
 1019 RPNKCPKGGGLAAVSTSVNLTSPGOVLASRFMAVHKLKNSODYTEALRAARLANITA 1078  
 1079 DLKRVGCTBPAREVFPYTTTNVFEYOYTLTBEGFLMLSLCYPTPAVSGLLGLDRSG 1138  
 1079 DLKRVGCTBPAREVFPYTTTNVFEYOYTLTBEGFLMLSLCYPTPAVSGLLGLDRSG 1138  
 1080 DLKRVGCTBPAREVFPYTTTNVFEYOYTLTBEGFLMLSLCYPTPAVSGLLGLDRSG 1139  
 1080 DLKRVGCTBPAREVFPYTTTNVFEYOYTLTBEGFLMLSLCYPTPAVSGLLGLDRSG 1139  
 1139 LNLTLISIMLVTVTFGFMALMDSYNAVSLINIVSAVGMVSEFVSHITRSFAISTKPTWL 1198  
 1139 LNLTLISIMLVTVTFGFMALMDSYNAVSLINIVSAVGMVSEFVSHITRSFAISTKPTWL 1198  
 1140 LNLTLISIMLVTVTFGFMALMDSYNAVSLINIVSAVGMVSEFVSHITRSFAISTKPTWL 1199  
 1140 LNLTLISIMLVTVTFGFMALMDSYNAVSLINIVSAVGMVSEFVSHITRSFAISTKPTWL 1199  
 1199 ERAKEATISGSAVPAVANTNLPGLIVGLAQAOLIQIFFRMLNLTLLGLHGLVFL 1258  
 1199 ERAKEATISGSAVPAVANTNLPGLIVGLAQAOLIQIFFRMLNLTLLGLHGLVFL 1258  
 1200 ERAKATVFMGSAVPAVANTNLPGLIVGLAQAOLIQIFFRMLNLTLLGLHGLVFL 1259  
 1200 ERAKATVFMGSAVPAVANTNLPGLIVGLAQAOLIQIFFRMLNLTLLGLHGLVFL 1259  
 1259 PVLSTVGPVPPVNALALBOKRABEAVAAVMAVSCPNHPSVSTADNITYNHSEBGS -KG 1317  
 1259 PVLSTVGPVPPVNALALBOKRABEAVAAVMAVSCPNHPSVSTADNITYNHSEBGS -KG 1317  
 1260 PVLSTVGPVPPVNALALBOKRABEAVAAVMAVSCPNHPSVSTADNITYNHSEBGS -KG 1316  
 1260 PVLSTVGPVPPVNALALBOKRABEAVAAVMAVSCPNHPSVSTADNITYNHSEBGS -KG 1316  
 1318 AGAISNPLPNNGROF 1332  
 1318 AGAISNPLPNNGROF 1332  
 1317 ANAARSSLPKSDOKF 1331  
 1317 ANAARSSLPKSDOKF 1331

RESULT 15  
 US-10-621-758A-12  
 ; Sequence 12: Application US/10621758A  
 ; Publication No. US20040093622A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Altman, Scott W  
 ; APPLICANT: Wang, Luquan

APPLICANT: Graziano, Michael  
 APPLICANT: Murgolo, Nick  
 TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF  
 FILE REFERENCE: JB01603-K-US  
 CURRENT APPLICATION NUMBER: US/10/621,758A  
 PRIOR FILING DATE: 2003-07-17  
 PRIOR APPLICATION NUMBER: 60/397,442  
 NUMBER OF SEQ ID NOS: 50  
 SOFTWARE: Patentin version 3.1  
 SEQ ID NO 12  
 LENGTH: 1333  
 TYPE: PRT  
 ORGANISM: Mus sp.  
 US-10-621-758A-12

Query Match 78.3%; Score 5407; DB 4; Length 1333;  
 Best Local Similarity 77.2%; Pred. No. 0;  
 Matches 1028; Conservative 126; Mismatches 174; Indels 4; Gaps 4;

4 AGLRGLWMLALLRLAQSPPYTTTHQPGYCAPYDECGKNPELGSIMTLNVSCLSNTPA 63  
 3 AAMQGLWMLALLRLNLSAQGLYPTTHKAGCTFYEBCGKNPELGSIMTLNVSCLSNTPA 62  
 64 RKTGDHLILQXICPRLYTPGN-10ACCSAQVLSLEASLITTKALLTRCPACSDNFVN 122  
 63 RHVTGDHLILQVRCPLVNGPNDYACCTQVLSLDSLSTTKALLTRCPACSDNFVS 122  
 123 LHCHNTCSPNOSLFINVTBVAOLAGQLPAVVAVEAFYHSPAFOSYDSCSRVPAAT 182  
 123 LHCHNTCSPNOSLFINVTBVAOLAGQLPAVVAVEAFYHSPAFOSYDSCSRVPAAT 182  
 183 LAVGTCGVGYSALCAQRMNLFQSDTGNGLAFLDITPHLBEQAVSGSIOPLNGVAR 242  
 183 LAVGTCGVGYSALCAQRMNLFQSDTGNGLAFLDITPHLBEQAVSGSIOPLNGVAR 242  
 243 CNEGQBDVATGSCQDCAASCPALAPOLDSFTFYIGMPGSLVLIILCSAVAVTILL 302  
 243 CNEGQBDVATGSCQDCAASCPALAPOLDSFTFYIGMPGSLVLIILCSAVAVTILL 302  
 243 CNEGQBDVATGSCQDCAASCPALAPOLDSFTFYIGMPGSLVLIILCSAVAVTILL 302  
 243 CNEGQBDVATGSCQDCAASCPALAPOLDSFTFYIGMPGSLVLIILCSAVAVTILL 302  
 303 VGRFVAPARDKSMVNDPKGTSLSDKLSFTHTLIGOPQSGMTVYASNPFLTVLSVIP 362  
 303 VGRFVAPARDKSMVNDPKGTSLSDKLSFTHTLIGOPQSGMTVYASNPFLTVLSVIP 362  
 303 VGRFVAPARDKSMVNDPKGTSLSDKLSFTHTLIGOPQSGMTVYASNPFLTVLSVIP 362  
 303 VGRFVAPARDKSMVNDPKGTSLSDKLSFTHTLIGOPQSGMTVYASNPFLTVLSVIP 362  
 363 VVALAGLVFTLTLTPVLEMSAPNSQARSSEKAFHDOHGFPPRTNOVLTAPNRSYRY 422  
 363 VVALAGLVFTLTLTPVLEMSAPNSQARSSEKAFHDOHGFPPRTNOVLTAPNRSYRY 422  
 363 VVALAGLVFTLTLTPVLEMSAPNSQARSSEKAFHDOHGFPPRTNOVLTAPNRSYRY 422  
 363 VVALAGLVFTLTLTPVLEMSAPNSQARSSEKAFHDOHGFPPRTNOVLTAPNRSYRY 422  
 423 DSIILGPKNPSGILDLLELELELOERLRLHLOVMSPEAORNI SLODI CYAPLNPNTSL 482  
 423 DSIILGPKNPSGILDLLELELELOERLRLHLOVMSPEAORNI SLODI CYAPLNPNTSL 482  
 483 YDCCISLLOYPONNRTLLLTANQTLMGOTSOVDKDHLYCANMPLFKDSTALAASC 542  
 483 YDCCISLLOYPONNRTLLLTANQTLMGOTSOVDKDHLYCANMPLFKDSTALAASC 542  
 483 YDCCISLLOYPONNRTLLLTANQTLMGOTSOVDKDHLYCANMPLFKDSTALAASC 542  
 483 YDCCISLLOYPONNRTLLLTANQTLMGOTSOVDKDHLYCANMPLFKDSTALAASC 542  
 543 MADYGAPEPFLAIGGYKDYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLERMA 602  
 543 MADYGAPEPFLAIGGYKDYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLERMA 602  
 543 MADYGAPEPFLAIGGYKDYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLERMA 602  
 543 MADYGAPEPFLAIGGYKDYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLERMA 602  
 603 FORSTADKFOIARSARSLEDEINRTTIQDLPPFAVSIIIVFLYISIALGSSYSWSRVAV 662  
 603 FORSTADKFOIARSARSLEDEINRTTIQDLPPFAVSIIIVFLYISIALGSSYSWSRVAV 662  
 603 FORSTADKFOIARSARSLEDEINRTTIQDLPPFAVSIIIVFLYISIALGSSYSWSRVAV 662  
 603 FORSTADKFOIARSARSLEDEINRTTIQDLPPFAVSIIIVFLYISIALGSSYSWSRVAV 662  
 663 DSKATIGLGAVAVVLAANMAAMGFYLGIRSSLVLIQVPPFLVLSGADNITIFVLEYQ 722  
 663 DSKATIGLGAVAVVLAANMAAMGFYLGIRSSLVLIQVPPFLVLSGADNITIFVLEYQ 722  
 663 DSKATIGLGAVAVVLAANMAAMGFYLGIRSSLVLIQVPPFLVLSGADNITIFVLEYQ 722  
 663 DSKATIGLGAVAVVLAANMAAMGFYLGIRSSLVLIQVPPFLVLSGADNITIFVLEYQ 722  
 723 RLPRRGEPEPEVHIGRALGRVAPSMILCSLSBAICFPLGALTPMPAVRTALSGLAIVL 782  
 723 RLPRRGEPEPEVHIGRALGRVAPSMILCSLSBAICFPLGALTPMPAVRTALSGLAIVL 782  
 723 RLPRRGEPEPEVHIGRALGRVAPSMILCSLSBAICFPLGALTPMPAVRTALSGLAIVL 782  
 723 RLPRRGEPEPEVHIGRALGRVAPSMILCSLSBAICFPLGALTPMPAVRTALSGLAIVL 782  
 783 DFLQMSAFVALSLDSKROEASRLDYCCVKEQOELPPPCQGBGLLGFPOKAYAPFLH 842  
 783 DFLQMSAFVALSLDSKROEASRLDYCCVKEQOELPPPCQGBGLLGFPOKAYAPFLH 842

```

Db      783 DFLQMTAFVALLSLDSKROEASRPDYVCCFSSRNLPPEKQKEGLLCCFRKIYTPFLH 842
Qy      843 WITRGVLLFLFALFVSVLSYSMCHISVGLDOELAPKDSYLDYFLFLNRYFEGAPYF 902
Db      843 RIRPVPVLLFLVLPFANILYLMCNISVGLDQDALPKDSYLDYFLFLNRYLEVGPYF 902
Qy      903 VTTLGNFSSBAQMAICSSAGCNFSFTQKIQYATEPPEQSYLAI PASSWDDFTDWT 962
Db      903 DTSGYNFSTAGMNAICSSAGCESFSLQKIQYASEFPNQSVAIAAGSWDDFTDWT 962
Qy      963 P-SSCCRLYISGPNKXPCPSPTNSLNCMKMSITMGSVRPSVEQFHXYLPWFINDRPN 1021
Db      963 PSSSCCRITRGRHNDPCPSDTSFNCKMCKNRTLGPVRPTTEQFHXYLPWFINDRPN 1022
Qy      1022 IKCPKGLAAYSTSVNLTSQVLA SRFMA YHKPLKNSODYTEALRARELAANITADLR 1081
Db      1023 IKCPKGLAAYRTSVNLSSDQI IASQFMA YHKPLKNSODFTBALPASRLAANITAEIR 1082
Qy      1082 KVPGETDPAEVPPYTTTNNFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLIN 1141
Db      1083 KVPGETDPNFEVPPYTTISNVFYQYLTVLPEGLFTLALCFVPPFVVCYLLGLDIRSGILN 1142
Qy      1142 LLSIVMLVDTYGFMA LMDISYRAVSLINLVSAVGSVEFVSHITRSFAISTKPTLERA 1201
Db      1143 LLSIIMILVDTIGLMAVWGISYNAVSLINLVTAVGSVEFVSHITRSFAVSTKPTLERA 1202
Qy      1202 KEATISMSGSAVPAGVAMTNLPGLVLGLAKAQLIQIFPRNLNLTLLGLHGLVFLPYI 1261
Db      1203 KDAITFMGSAVPAVAMTNFPGLILIGFAQAOIQIFPRNLNLTLLGLHGLVFLPVY 1262
Qy      1262 LSYVGPDPALALEQKRAEBAVAAMVASCPNHPRSVSTADNIYVNHSPESG-IKGAQA 1320
Db      1263 LSYLGPDPVQALVLEBKATEA-AMVSEBSCPOYPPADANTSDDYVNGFNPFIPEINA 1321
Qy      1321 ISNFLPNNGROF 1332
Db      1322 ASSSLPKSDQKF 1333

```

Search completed: March 22, 2006, 23:10:54  
 Job time : 194 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 22:47:02 ; Search time 201 Seconds

(without alignments)  
2911.704 Million cell updates/sec

Title: US-10-736-769-4

Sequence: 1 MEAGRGRLMALTLRLAQ.....GSIKGAGISNPLPNNGRQF 1332

Scoring table:

BLOSUM62

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6909	100.0	1332	ADJ27297	Adj27297 Human NPC
2	6909	100.0	1332	ADY60740	Ady60740 Human NPC
3	6909	100.0	1332	ABE93570	ABE93570 Human NPC
4	6896	99.8	1332	AAW79169	AAW79169 Human pro
5	6896	99.8	1332	AAW79169	AAW79169 Human pro
6	6872.5	99.5	1359	AAW79168	AAW79168 Human pro
7	6872.5	99.5	1359	ADJ27337	ADJ27337 Human NPC
8	6872.5	99.5	1359	ADY60780	ADY60780 Human NPC
9	6872.5	99.5	1359	ABE93570	ABE93570 Human NPC
10	6536	94.6	1344	ABG22693	ABG22693 Novel hum
11	5421.5	78.5	1331	ADJ27295	ADJ27295 Rat NPCIL
12	5421.5	78.5	1331	ADY60738	ADY60738 Rat Niema
13	5421.5	78.5	1331	ABE93568	ABE93568 Rat NPCIL
14	5407	78.3	1333	ADJ27305	ADJ27305 Mouse NPC
15	5407	78.3	1333	ADY60748	ADY60748 Mouse NPC
16	5407	78.3	1333	ABE93578	ABE93578 Mouse NPC
17	4466	64.6	982	ABG22691	ABG22691 Novel hum
18	2402.5	34.8	1278	AAW88445	AAW88445 Human NPC
19	2402.5	34.8	1278	AAW88445	AAW88445 Human NPC
20	2402.5	34.8	1278	ADQ39879	ADQ39879 Human myo
21	2402.5	34.8	1278	ADU06723	ADU06723 Human bro
22	2385	34.5	1319	AAW88446	AAW88446 Mouse NPC
23	2262	32.7	1287	ABE61737	ABE61737 Drosophil
24	2262	32.7	1287	ADG96670	ADG96670 Drosophil

25	1864.5	27.0	1223	4	ABE58629	ABE58629 Drosophil
26	1329	19.2	1170	2	AAW88447	AAW88447 Yeast NPC
27	1078	15.6	1296	2	AAW88448	AAW88448 Caenorhab
28	963	13.9	1229	4	AAW80153	AAW80153 Human pro
29	963	13.9	1229	4	AAW80152	AAW80152 Human pro
30	942	13.6	1294	4	AAW5637	AAW5637 SFP domai
31	651.5	9.4	1447	2	AAW5375	AAW5375 Human pat
32	651.5	9.4	1447	2	AAW52200	AAW52200 Human pat
33	651.5	9.4	1447	2	AAW72969	AAW72969 Human pat
34	651.5	9.4	1447	4	AAW67163	AAW67163 Human pat
35	651.5	9.4	1447	5	AAW19830	AAW19830 Human pat
36	651.5	9.4	1447	5	ABJ10931	ABJ10931 TRC8 rela
37	651.5	9.4	1447	5	AAW79571	AAW79571 Human pat
38	651.5	9.4	1447	7	ABU62275	ABU62275 Human pat
39	651.5	9.4	1447	7	ADU6678	ADU6678 Human pro
40	651.5	9.4	1447	7	ADU6678	ADU6678 Human pro
41	651.5	9.4	1447	7	ADU6678	ADU6678 Human pro
42	651.5	9.4	1447	8	ADU6678	ADU6678 Human pro
43	646.5	9.4	1434	2	AAW52199	AAW52199 Human pat
44	646.5	9.4	1434	2	AAW52199	AAW52199 Human pat
45	646.5	9.4	1434	4	AAW67159	AAW67159 Murine pa

## ALIGNMENTS

## RESULT 1

ADJ27297 standard; protein: 1332 AA.

ADJ27297;

20-MAY-2004 (first entry)

Human NPCIL.

Niemann-Pick disease; type C1; gene-like 1; NPCIL1; trans-golgi network;

plasma membrane; transport signal; promoter;

sterol regulated element binding protein 1; SREBP1;

binding consensus sequence; transmembrane domain; sterol-sensing domain;

SSD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;

cholesterol absorption; serum cholesterol; hyperlipidaemia;

atherosclerosis; coronary heart disease; stroke; arteriosclerosis.

Homo sapiens.

WO2004009772-A2.

29-JAN-2004.

17-JUL-2003; 2003WO-US022467.

19-JUL-2002; 2002US-0397442P.

(SCHE ) SCHERING CORP.

Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;

WPI; 2004-132945/13.

N-PSDB; ADJ27296.

GENBANK; AF192522.

New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting

or identifying antagonists of NPCIL1 for inhibiting intestinal

cholesterol absorption in a subject, or for treating elevated serum

cholesterol or stroke.

Claim 11; SEQ ID NO 4; 125bp; English.

This sequence represents a Niemann-Pick disease, type C1, gene-like 1 (NPCIL1) polypeptide. NPCIL1 is an N-glycosylated protein which contains a motif ADJ27331 which acts as a trans-golgi network to plasma membrane transport signal, and which exhibits limited tissue distribution and

CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a  
 CC sterol regulated element binding protein 1 (SREBP1) binding consensus  
 CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-  
 CC sensing domain (SSD) which is involved in sensing cholesterol levels,  
 CC possibly by a mechanism which involves direct cholesterol binding. NPC1L1  
 CC has 42% amino acid sequence homology to human NPC1, a receptor  
 CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or  
 CC polynucleotides are useful for detecting or identifying antagonists of  
 CC NPC1L1, which can be used for inhibiting intestinal cholesterol  
 CC absorption in a subject, or for treating medical conditions including  
 CC elevated serum cholesterol, hyperlipidaemia, atherosclerosis, coronary  
 CC heart disease, stroke or arteriosclerosis.

XX Sequence 1332 AA;

Query Match 100.0%; Score 6909; DB 8; Length 1332;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEAGLRGWLMLALRLAQSSEPTTIHQGYCAFYDECGKNEPESGSLMTLSNVSLSN 60  
 DB 1 MAEAGLRGWLMLALRLAQSSEPTTIHQGYCAFYDECGKNEPESGSLMTLSNVSLSN 60  
 QY 61 TPARKITGSHLLLOKICPRLYTGPRTOACCSAKOLVSLASITKALLTRCPACSDNF 120  
 DB 61 TPARKITGSHLLLOKICPRLYTGPRTOACCSAKOLVSLASITKALLTRCPACSDNF 120  
 QY 121 VNLHCNNTSPNOSTLTINTVAVQLAGOLPAVVAEAFYQHSFAEOSYDSCSRVPPAA 180  
 DB 121 VNLHCNNTSPNOSTLTINTVAVQLAGOLPAVVAEAFYQHSFAEOSYDSCSRVPPAA 180  
 QY 121 VNLHCNNTSPNOSTLTINTVAVQLAGOLPAVVAEAFYQHSFAEOSYDSCSRVPPAA 180  
 DB 121 VNLHCNNTSPNOSTLTINTVAVQLAGOLPAVVAEAFYQHSFAEOSYDSCSRVPPAA 180  
 QY 181 ATLAVGTCGCVGSALCNARWLNFOQDTGNGLAPLDITFHLLPEQAVSGGIQPINEG 240  
 DB 181 ATLAVGTCGCVGSALCNARWLNFOQDTGNGLAPLDITFHLLPEQAVSGGIQPINEG 240  
 QY 241 AECNESQSDVATCSCODCAASCPAARPAOLDSTVYLGOMPSVLIIILCSVPVWVT 300  
 DB 241 AECNESQSDVATCSCODCAASCPAARPAOLDSTVYLGOMPSVLIIILCSVPVWVT 300  
 QY 301 LLVGFPAVARADSKKAVDPKKGTSLSDKLSFTHTLLGQFGOMGTWVSWPPLTILVLS 360  
 DB 301 LLVGFPAVARADSKKAVDPKKGTSLSDKLSFTHTLLGQFGOMGTWVSWPPLTILVLS 360  
 QY 361 IPVVALAAGLVFTELTTPDVELMSAPNSQARSEKAFHDQFGFPRTNOYLTAAPRSSY 420  
 DB 361 IPVVALAAGLVFTELTTPDVELMSAPNSQARSEKAFHDQFGFPRTNOYLTAAPRSSY 420  
 QY 421 RYDSILLGRKNSFGIILDLLELLELQERLRLQVWSPBAQNISLQDI CVAPLPNDT 480  
 DB 421 RYDSILLGRKNSFGIILDLLELLELQERLRLQVWSPBAQNISLQDI CVAPLPNDT 480  
 QY 481 SLVDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKQGTALAL 540  
 DB 481 SLVDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKQGTALAL 540  
 QY 541 SCWADYGAVPFPFLAIGYKXGKDYSEAEALIMTFSLNTPYAGDPRLAQAALMEAEFL 600  
 DB 541 SCWADYGAVPFPFLAIGYKXGKDYSEAEALIMTFSLNTPYAGDPRLAQAALMEAEFL 600  
 QY 601 RAFORMAAGFOVTFPAERSLDEINRTTAEDLPITATSIYIVLYTSLALGYSMSRY 660  
 DB 601 RAFORMAAGFOVTFPAERSLDEINRTTAEDLPITATSIYIVLYTSLALGYSMSRY 660  
 QY 661 MYDSKATLGLGAVAVILGAVMAAMGFYSYIGRSSLYILOVPELVLSVAGADNIFIVLE 720  
 DB 661 MYDSKATLGLGAVAVILGAVMAAMGFYSYIGRSSLYILOVPELVLSVAGADNIFIVLE 720  
 QY 721 YORLPRRPGEPREVIHGRALGRVAPSWLCSLEBALCFPLGALTMPPAVTFALTGLAY 780  
 DB 721 YORLPRRPGEPREVIHGRALGRVAPSWLCSLEBALCFPLGALTMPPAVTFALTGLAY 780  
 QY 781 IIDFLLQMSAFVALSLDSKROBASRLDVCCYKPOELPPGGEGILLGFPQKAYAPFL 840  
 DB 781 IIDFLLQMSAFVALSLDSKROBASRLDVCCYKPOELPPGGEGILLGFPQKAYAPFL 840

DB 781 IIDFLLQMSAFVALSLDSKROBASRLDVCCYKPOELPPGGEGILLGFPQKAYAPFL 840  
 QY 841 LHMITRGVLLFLALFGVSLVSMCHISVGLDQELAPKDSYLLDYFLNRYFEVGAPY 900  
 DB 841 LHMITRGVLLFLALFGVSLVSMCHISVGLDQELAPKDSYLLDYFLNRYFEVGAPY 900  
 QY 901 YPPTTIGYVPSBAGNNAICSSAGCNBSFTQKIQTATFPBPBSYLAIPASSVVDPIWM 960  
 DB 901 YPPTTIGYVPSBAGNNAICSSAGCNBSFTQKIQTATFPBPBSYLAIPASSVVDPIWM 960  
 QY 961 LTPSSCCRLYISGPNKDKCPSTVNSLNCIKCMSTTMSVPRSVQFHKYLPWFLNDRP 1020  
 DB 961 LTPSSCCRLYISGPNKDKCPSTVNSLNCIKCMSTTMSVPRSVQFHKYLPWFLNDRP 1020  
 QY 1021 NIKCPKGLAAYSTSVNLTSQVLA SRMAVYHKPKNSQDYTEALRAARELAANTADL 1080  
 DB 1021 NIKCPKGLAAYSTSVNLTSQVLA SRMAVYHKPKNSQDYTEALRAARELAANTADL 1080  
 QY 1081 RKVPGTDPAEVFPPTITVTFYQYLTILPEGLFMLSCLVPTFAVSCILLGLDLSGLL 1140  
 DB 1081 RKVPGTDPAEVFPPTITVTFYQYLTILPEGLFMLSCLVPTFAVSCILLGLDLSGLL 1140  
 QY 1141 NLSITVMIIVDTYGFMAAMDISTYNAVSLINLVSAVMSVEFYSHITRSFAISTKPTWLR 1200  
 DB 1141 NLSITVMIIVDTYGFMAAMDISTYNAVSLINLVSAVMSVEFYSHITRSFAISTKPTWLR 1200  
 QY 1201 AKEATISMSGSAVPAVAMTNLPGLIVLGLAKQLIQIPFFRLNLTLLGLHLGLFLPY 1260  
 DB 1201 AKEATISMSGSAVPAVAMTNLPGLIVLGLAKQLIQIPFFRLNLTLLGLHLGLFLPY 1260  
 QY 1261 ILSYGPDPVNPALALEQKAEBAVAAVMAASCPNHSRSTADNITYVNSFESITGAGA 1320  
 DB 1261 ILSYGPDPVNPALALEQKAEBAVAAVMAASCPNHSRSTADNITYVNSFESITGAGA 1320  
 QY 1321 ISNFLPNNGROF 1332  
 DB 1321 ISNFLPNNGROF 1332

RESULT 2  
 ID ADY60740 standard; protein; 1332 AA.  
 XX ADY60740;  
 AC 19-MAY-2005 (first entry)  
 DT XX  
 DE Human Niemann-Pick C1-like protein 1 (NPC1L1) protein SEQ ID 4.  
 XX  
 KW Niemann-Pick C1-like protein 1; Antilipemic; Antiarteriosclerotic;  
 KW Cardiac; Vascular; Cerebroprotective; cholesterol; hyperlipidemia;  
 KW metabolic disorder; atherosclerosis; cardiovascular disease;  
 KW coronary artery disease; cerebrovascular ischemia; neurological disease;  
 KW arteriosclerosis; transgenic animal.  
 XX  
 OS Homo sapiens.  
 OS XX  
 PN MO2005015988-AL.  
 XX  
 PD 24-FEB-2005.  
 XX  
 PF 16-DEC-2003; 2003WO-US040113.  
 XX  
 PR 17-JUL-2003; 2003US-00621758.  
 PR 22-AUG-2003; 2003US-00646301.  
 PR 16-SEP-2003; 2003US-00663208.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Altman SW, Murgolo NJ, Wang LQ, Graziano MP;  
 XX  
 DR WPI; 2005-284403/29.  
 DR N-PSDB; ADY60739, ADY60787.



DE Human NPC1L1 protein.  
KW Protein engineering; NPC1L1; antidiabetic; antiarteriosclerotic;  
KW cardiant; vasotropic; hyperlipidemia; atherosclerosis; stroke;  
KW arteriosclerosis; human.  
XX Homo sapiens.  
PN WO2005069900-A2.  
XX 04-AUG-2005.  
PD  
PF 14-JAN-2005; 2005MO-US001469.  
PR 16-JAN-2004; 2004US-0537341P.  
PA (MERI ) MERCK & CO INC.  
XX Garcia-Calvo M;  
PI WPI; 2005-564070/57.  
DR N-PSDB; AEB93569.  
XX  
XX  
PT Identifying ligand of NPC1L1 for stimulating the activity of NPC1L1, by  
PT contacting NPC1L1 with detectably labeled substituted 2-acetidinone  
PT glucuronide and a candidate compound and determining if compound binds to  
PT human NPC1L1.  
PS Example 2; SEQ ID NO 4; 215pp; English.  
XX  
XX The invention relates to identifying a ligand of NPC1L1. The method  
CC involves contacting human NPC1L1 with a detectably labeled substituted 2-  
CC acetidinone glucuronide and a candidate compound and determining whether  
CC the candidate compound binds to human NPC1L1. In identifying a ligand of  
CC NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10  
CC nM or lower. The detectably labeled substituted 2-acetidinone glucuronide  
CC is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands  
CC are useful for stimulating or blocking the activity of NPC1L1, and for  
CC treating conditions caused or mediated by NPC1L1. It is useful for  
CC reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart  
CC disease, stroke, or arteriosclerosis. The present sequence represents a  
CC human NPC1L1, a N-glycosylated protein.  
XX  
SQ Sequence 1332 AA;  
Query Match 100.0%; Score 6909; DB 9; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 IPVALLAGLVFTELTTDPVELMSAPNSQARSEKA FHOHGGPFRTNOVILLTAPRBSY 420  
DB 361 IPVALLAGLVFTELTTDPVELMSAPNSQARSEKAPHDHGPFRTNOVILLTAPRBSY 420  
QY 421 RYDSLLGPKNSGILDDLLELLELOERLHLQVWSEARNTSLDICYAPLPNDT 480  
DB 421 RYDSLLGPKNSGILDDLLELLELOERLHLQVWSEARNTSLDICYAPLPNDT 480  
QY 481 SLYDCCINSLLQYFONNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLYTKDGTALAL 540  
DB 481 SLYDCCINSLLQYFONNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLYTKDGTALAL 540  
QY 541 SCMAHYGAPVPEFLAIGYKGDYSEAEALIMTFSLNNYPADDPRLAOKLMEEAFLBEM 600  
DB 541 SCMAHYGAPVPEFLAIGYKGDYSEAEALIMTFSLNNYPADDPRLAOKLMEEAFLBEM 600  
QY 601 RAFORRMAGMFOVTFTAERSLEDEINRTTAEDLPITATSYYIPLYISLALGSYSWSRY 660  
DB 601 RAFORRMAGMFOVTFTAERSLEDEINRTTAEDLPITATSYYIPLYISLALGSYSWSRY 660  
QY 661 WYDSKATLGLGGVAVVLGA VMAAMGPFYSYLGIRSSLVILQVVPFLVLSVADNIFI FVLE 720  
DB 661 WYDSKATLGLGGVAVVLGA VMAAMGPFYSYLGIRSSLVILQVVPFLVLSVADNIFI FVLE 720  
QY 721 YQRLRRRREPREVHITGALGRVARSMLCSISEALCFPLGALTMPANRTFALSGLA 780  
DB 721 YQRLRRRREPREVHITGALGRVARSMLCSISEALCFPLGALTMPANRTFALSGLA 780  
QY 781 IIDPFLQMSAFVALLSLDSKROEASRLDVCCVCKQELPPPEQSGELLLGFFQKAYAPRL 840  
DB 781 IIDPFLQMSAFVALLSLDSKROEASRLDVCCVCKQELPPPEQSGELLLGFFQKAYAPRL 840  
QY 841 LHMTRGVVLLFLALFGVSLYSWCHISVGLDQELALPKDYLDPFLPLNRYFEVGA 900  
DB 841 LHMTRGVVLLFLALFGVSLYSWCHISVGLDQELALPKDYLDPFLPLNRYFEVGA 900  
QY 901 YVTTLLGVNSESSEAMNMCSSAGCNNSFPQKIYATEFPQSYLALPASSVDDFDW 960  
DB 901 YVTTLLGVNSESSEAMNMCSSAGCNNSFPQKIYATEFPQSYLALPASSVDDFDW 960  
QY 961 LTPSSCCRXYISGPKNDKFCPESTVNSLCKKCNKSIITGSAVSPVEQPHKYLPMFLNDRP 1020  
DB 961 LTPSSCCRXYISGPKNDKFCPESTVNSLCKKCNKSIITGSAVSPVEQPHKYLPMFLNDRP 1020  
QY 1021 NIKCPKGLAAYSTSVNLTSQVTLASRPMAVHKPLKNSQDYTEALRAARELAANITADL 1080  
DB 1021 NIKCPKGLAAYSTSVNLTSQVTLASRPMAVHKPLKNSQDYTEALRAARELAANITADL 1080  
QY 1081 RRVPGTDAFEVFPYTTITNVFEQYLTLLPBEGLFMLSCLVPTFAVSCULLGLDRSGLL 1140  
DB 1081 RRVPGTDAFEVFPYTTITNVFEQYLTLLPBEGLFMLSCLVPTFAVSCULLGLDRSGLL 1140  
QY 1141 NLLSTVIMLVDTVGFPMALMDISYNVSLINLVSAAGMSVEFSSHITRSFAISTKPTWLER 1200  
DB 1141 NLLSTVIMLVDTVGFPMALMDISYNVSLINLVSAAGMSVEFSSHITRSFAISTKPTWLER 1200  
QY 1201 AKEATISMGSAVFAVAMTNLPGLIIVLGLAKQLIQIEFFRLNLLITLLGLHGLVFLPV 1260  
DB 1201 AKEATISMGSAVFAVAMTNLPGLIIVLGLAKQLIQIEFFRLNLLITLLGLHGLVFLPV 1260  
QY 1261 IISYVGPVNPALALEQKRAEBAVAAMVASCPNHPRSVSTADNITYNHSFEGSIKGA 1320  
DB 1261 IISYVGPVNPALALEQKRAEBAVAAMVASCPNHPRSVSTADNITYNHSFEGSIKGA 1320  
QY 1321 ISNFLPNNGROP 1332  
DB 1321 ISNFLPNNGROP 1332  
RESULT 4  
AAM79169  
ID AAM79169 standard; protein; 1332 AA.  
XX



AC	AAW79169;	
XX		
DT	06-NOV-2001 (first entry)	
XX		
DE	Human protein SEQ ID NO 1831.	
XX		
KW	Human: cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	nervous system disorder; arthritis; inflammation.	
XX		
XX	Homo sapiens.	
XX		
PN	WO200157190-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	05-FEB-2001; 2001WO-US004098.	
XX		
PR	03-FEB-2000; 2000US-00496914.	
PR	27-APR-2000; 2000US-00560875.	
PR	20-JUN-2000; 2000US-00598075.	
PR	19-JUL-2000; 2000US-00620325.	
PR	01-SEP-2000; 2000US-00654936.	
PR	15-SEP-2000; 2000US-00663561.	
PR	20-OCT-2000; 2000US-00693325.	
PR	30-NOV-2000; 2000US-00728422.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;	
PI	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;	
PI	Xue AJ, Yang Y, Wejhtman T, Goodrich R;	
XX		
DR	WPI: 2001-476283/51.	
DR	N-PSDB: AAK52302.	
PT	Nucleic acids encoding polypeptides with cytokine-like activities, useful	
PT	in diagnosis and gene therapy.	
XX		
PS	Claim 20; Page 4216-4219; 6221dp; English.	
XX		
XX	The invention relates to polynucleotides (AAK51456-AAK53435) and the	
CC	encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to	
CC	cytokine, cell proliferation or cell differentiation or which may induce	
CC	production of other cytokines in other cell populations. The	
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or	
CC	peptide therapy. The polypeptides have various cytokine-like activities,	
CC	e.g. stem cell growth factor activity, haematopoiesis regulating	
CC	activity, tissue growth factor activity, immunomodulatory activity and	
CC	activin/inhibin activity and may be useful in the diagnosis and/or	
CC	treatment of cancer. Leukaemia, nervous system disorders, arthritis and	
CC	inflammation. Note: Records for SEQ ID NO 2110 (AAK52881), 2111	
CC	(AAK52882) and 3666 (AAW80020) are omitted as the relevant pages from the	
CC	sequence listing were missing at the time of publication	
XX		
XX		
XX	Sequence 1332 AA;	
	Query Match 99.8%; Score 6896; DB 4; Length 1332;	
	Best Local Similarity 99.8%; Pred. No. 0;	
	Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0	
OY	1 MAEAGLRMLLWALLRLAASEPYTHIQPGYCAFYDCGKNPELSGSLMTLSNVSCLSN 60	
DB	1 MAEAGLRMLLWALLRLAASEPYTHIQPGYCAFYDCGKNPELSGSLMTLSNVSCLSN 60	
OY	61 TPARKITGDHLLLOKICPRLYTGNTQACCSAKQLVSLBASLSTTKALLTRCPACSDNF 120	
DB	61 TPARKITGDHLLLOKICPRLYTGNTQACCSAKQLVSLBASLSTTKALLTRCPACSDNF 120	
OY	121 VNLHCNTGSPQSLFIVNTRVAQGAQQLPAVVAVYEAFFQHSFAEQSYDSCSRVRVPA 180	
DB	121 VNLHCNTGSPQSLFIVNTRVAQGAQQLPAVVAVYEAFFQHSFAEQSYDSCSRVRVPA 180	

QY	181	ATLAVGTCGYVYGSALCLNAQWNLNFGQSTGNGCLAPLDTTFHLLHSGQAVSGSIOPLNGBV	240
Db	181	ATLAVGTCGYVYGSALCLNAQWNLNFGQSTGNGCLAPLDTTFHLLHSGQAVSGSIOPLNGBV	240
QY	241	ARCNSQGDVATCSQDCCASCPAIPQALDSFFYLGOMGSLVLIIILCSYPAVVTI	300
Db	241	ARCNSQGDVATCSQDCCASCPAIPQALDSFFYLGOMGSLVLIIILCSYPAVVTI	300
QY	301	LLVGERVAPADKSKMVDPKKGTSLSDKLFSFTHTLTGQPFQMGSTWASWPLTLLVLSV	360
Db	301	LLVGERVAPADKSKMVDPKKGTSLSDKLFSFTHTLTGQPFQMGSTWASWPLTLLVLSV	360
QY	361	IPVVALAAGLYFTLTTDPVBLMSAPNQASSEKAFHOHGFERTNOVILTAPNRSY	420
Db	361	IPVVALAAGLYFTLTTDPVBLMSAPNQASSEKAFHOHGFERTNOVILTAPNRSY	420
QY	421	RYDSLILGPKNFSGLIDDLLELELELOERLRLHQLQVMSPEAQNRSLSDICYPALNPDT	480
Db	421	RYDSLILGPKNFSGLIDDLLELELELOERLRLHQLQVMSPEAQNRSLSDICYPALNPDT	480
QY	481	SLYDCCINSLLQYFONNFTLLLTANQTLMGQTSQVMDKHFLYCANAPLTFKQGTALAL	540
Db	481	SLYDCCINSLLQYFONNFTLLLTANQTLMGQTSQVMDKHFLYCANAPLTFKQGTALAL	540
QY	541	SCMDYGPVPEPLAIGYKGDYSEARALIMTFSLINYPAGDPLLAQKLEBEAFLEEM	600
Db	541	SCMDYGPVPEPLAIGYKGDYSEARALIMTFSLINYPAGDPLLAQKLEBEAFLEEM	600
QY	601	RAFORRMAGMGOVPTAERSLEDEINRTTAEDPLFATSYIYIPLIYISLGSYSMSRY	660
Db	601	RAFORRMAGMGOVPTAERSLEDEINRTTAEDPLFATSYIYIPLIYISLGSYSMSRY	660
QY	661	MVDSKATIGLGGVAVYLGAWAAMGFBSYLGIRSSLVLQVWPPLVLSVGDANIITFVLE	720
Db	661	MVDSKATIGLGGVAVYLGAWAAMGFBSYLGIRSSLVLQVWPPLVLSVGDANIITFVLE	720
QY	721	YORLPRRGEPREVHIGALGRVAPBSMLCSSEALICEFPLGATMPAPRTALTSGLAV	780
Db	721	YORLPRRGEPREVHIGALGRVAPBSMLCSSEALICEFPLGATMPAPRTALTSGLAV	780
QY	781	ILDPLQMSAFVALSLSDSKRQOASRLDVCCKVQOBLPPQOGSGLLGFFQKAYAPFL	840
Db	781	ILDPLQMSAFVALSLSDSKRQOASRLDVCCKVQOBLPPQOGSGLLGFFQKAYAPFL	840
QY	841	LHWITRGVALLFLALFGVSLYSMCHISVGDQELALPKDSTILDYPLFLNRYFEVGAPV	900
Db	841	LHWITRGVALLFLALFGVSLYSMCHISVGDQELALPKDSTILDYPLFLNRYFEVGAPV	900
QY	901	YFVTTLIGNFSEBAGMNAICSSAGCNNSPFQKIOYATPEPQSLAIIPASSMVDPIEM	960
Db	901	YFVTTLIGNFSEBAGMNAICSSAGCNNSPFQKIOYATPEPQSLAIIPASSMVDPIEM	960
QY	961	LTPSSCCRLYISGPNKDFCPSTVNSLNCLEKNCMSITMGVSAPSYEQPHKYLPMFLNDRP	1020
Db	961	LTPSSCCRLYISGPNKDFCPSTVNSLNCLEKNCMSITMGVSAPSYEQPHKYLPMFLNDRP	1020
QY	1021	NIKCKKGLAAYSTSVNLTSQOVLASRFMAVHKRLKNSODYTEALRAARELANITAPL	1080
Db	1021	NIKCKKGLAAYSTSVNLTSQOVLASRFMAVHKRLKNSODYTEALRAARELANITAPL	1080
QY	1081	RKVPETDAFEVFPYTTINVFPEQYLLTLPBGLFMLSCLIVPTPAVSCULLGIDLRSGLL	1140
Db	1081	RKVPETDAFEVFPYTTINVFPEQYLLTLPBGLFMLSCLIVPTPAVSCULLGIDLRSGLL	1140
QY	1141	NLSTVIMLIVDTGFMALMDSISYNAVSLINVSANGMSVEFSSHITRSPALSTKPTWLER	1200
Db	1141	NLSTVIMLIVDTGFMALMDSISYNAVSLINVSANGMSVEFSSHITRSPALSTKPTWLER	1200
QY	1201	AKBATISGSAVPAVAMTNIPGLIWLGLAKAQLQIEFFRNLNLITLLGLHLHGVLPLV	1260
Db	1201	AKBATISGSAVPAVAMTNIPGLIWLGLAKAQLQIEFFRNLNLITLLGLHLHGVLPLV	1260

QY 1261 ILSYGPDPVNPALALEQKRAEENVAAMVASCNPHSRVSTADNITYVNSFEGSIKGAGA 1320  
DB 1261 ILSYGPDPVNPALALEQKRAEENVAAMVASCNPHSRVSTADNITYVNSFEGSIKGAGA 1320  
QY 1321 ISNFLPNNGROF 1332  
DB 1321 ISNFLPNNGROF 1332  
RESULT 5  
AAG65638  
ID AAG65638 standard; protein; 1332 AA.  
AC AAG65638;  
DT 07-JAN-2002 (first entry)  
DE SSD-containing SSP1 protein sequence.  
KW SSD; steroid-sensing domain; human; liver; testis; brain; cancer;  
KW neotropic; neuroprotective; antidiabetic; antiarteriosclerotic;  
KW cytostatic; antilipemic; SSP1.  
OS Homo sapiens.  
PN WO200170974-A1.  
PD 27-SEP-2001.  
PF 22-MAR-2001; 2001WO-JP002279.  
PR 24-MAR-2000; 2000JP-00088595.  
PA (TAKE ) TAKEDA CHEM IND LTD.  
PI Taniyama Y, Kita S, Komiya T;  
DR WPI; 2001-611501/70.  
DR N-PSDB; AA166914.  
XX  
XX New steroid-sensing domain-containing protein for diagnosing and screening  
PT candidate compounds in drug development for diabetes, obesity, cancer,  
PT arteriosclerosis, hyperlipidemia and neurodegenerative disorders.  
XX  
XX Claim 7; Page 115-122; 171pp; Japanese.  
PS  
CC The invention provides a novel SSD (steroid-sensing domain)-containing  
CC protein. The protein originates from human liver, human testis or human  
CC brain. The protein can be expressed by standard recombinant methodology.  
CC The proteins, encoded DNAs and antibodies are useful in diagnosis and  
CC screening candidate compounds in drug development for diabetes, obesity,  
CC cancer, arteriosclerosis, hyperlipemia, neurodegenerative disorders such  
CC as Alzheimer's disease and neural disorders. The present sequence  
CC represents the SSP1 protein which contains the SSD domain  
XX  
SQ Sequence 1332 AA;  
Query Match 99.8%; Score 6896; DB 4; Length 1332;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MBEAGLGMWLMVALRLAQLSEPTTIHQPGYCAFYDECGKPELGSIMTLISNVCLSN 60  
DB 1 MBEAGLGMWLMVALRLAQLSEPTTIHQPGYCAFYDECGKPELGSIMTLISNVCLSN 60  
QY 61 TPARKTGTGHLILQKICPRLTYGNTQACCSAKQVLSLEASISTKALLTRCPACSDNF 120  
DB 61 TPARKTGTGHLILQKICPRLTYGNTQACCSAKQVLSLEASISTKALLTRCPACSDNF 120  
QY 121 VNLHGNTGSPNOSLINTVTRVAQLGAGOLPAVVAAEAFYQHSFPAQSYDSCSRVAVPA 180  
DB 121 VNLHGNTGSPNOSLINTVTRVAQLGAGOLPAVVAAEAFYQHSFPAQSYDSCSRVAVPA 180

QY 181 ATLAVGTMGVYGSALCNARWLNQSDPTGNGLAIDITFHLLEFGQAVGSGIOPLNBSV 240  
DB 181 ATLAVGTMGVYGSALCNARWLNQSDPTGNGLAIDITFHLLEFGQAVGSGIOPLNBSV 240  
QY 241 ARCNESQGDVATCSCQPCASCPAIPARQALDSFTFYIGQMGSLVLIILICSVAVVTI 300  
DB 241 ARCNESQGDVATCSCQPCASCPAIPARQALDSFTFYIGQMGSLVLIILICSVAVVTI 300  
QY 301 ILVGRVVAARDKSKMVDPKKGTSLSDKLSFSTHTLLCOFQGMGTWVAWPLTLVLSV 360  
DB 301 ILVGRVVAARDKSKMVDPKKGTSLSDKLSFSTHTLLCOFQGMGTWVAWPLTLVLSV 360  
QY 361 IPVVALLAGLVFTELTTPDVELMSAPNSQABSEKAFHOHRSBPFRITNOVILTAANBSY 420  
DB 361 IPVVALLAGLVFTELTTPDVELMSAPNSQABSEKAFHOHRSBPFRITNOVILTAANBSY 420  
QY 421 RYDSLILGPKNFSGLIDLDLLELELELOBRRLHQLQVMSPEAQRNLSLODICYAPLNPN 480  
DB 421 RYDSLILGPKNFSGLIDLDLLELELELOBRRLHQLQVMSPEAQRNLSLODICYAPLNPN 480  
QY 481 SLVDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGTAAL 540  
DB 481 SLVDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGTAAL 540  
QY 541 SCMDVYGAPEVFFFLAIGYKGDYSEARALMTESLANTYPAGDPRLAQAQKMEAFLEEM 600  
DB 541 SCMDVYGAPEVFFFLAIGYKGDYSEARALMTESLANTYPAGDPRLAQAQKMEAFLEEM 600  
QY 601 RAFORRMAGMFOVTPAERSLEDEINRTTAEDLPFATSYIVIFLYISALGYSWSRV 660  
DB 601 RAFORRMAGMFOVTPAERSLEDEINRTTAEDLPFATSYIVIFLYISALGYSWSRV 660  
QY 661 MYDSKATLGLGVAVVLGAVMAAMGFYSYLGIRSSLVLQVVPFLVSVGADNIFLYLE 720  
DB 661 MYDSKATLGLGVAVVLGAVMAAMGFYSYLGIRSSLVLQVVPFLVSVGADNIFLYLE 720  
QY 721 YORLRRRGEPEBEVHIGALGRVAPSMILCSISEAICFFGLALTMPAVRTALNSGLAV 780  
DB 721 YORLRRRGEPEBEVHIGALGRVAPSMILCSISEAICFFGLALTMPAVRTALNSGLAV 780  
QY 781 IIDFLLQNSAFALLSLDSKROEASRLDVCCVKKQOEIPPGQGEGLLGFQKAYAPFL 840  
DB 781 IIDFLLQNSAFALLSLDSKROEASRLDVCCVKKQOEIPPGQGEGLLGFQKAYAPFL 840  
QY 841 LHMITRGVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAV 900  
DB 841 LHMITRGVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAV 900  
QY 901 YVVTLLGYNFSEAGMNAICSSAGCMNFSFTQIQYATBPPQSYLAI PASWVDFIDW 960  
DB 901 YVVTLLGYNFSEAGMNAICSSAGCMNFSFTQIQYATBPPQSYLAI PASWVDFIDW 960  
QY 961 LPPSSCRLYISGPKDKFCPTVNSLNLCKNCSITNGSVSPSEQFHKYLPFLNRP 1020  
DB 961 LPPSSCRLYISGPKDKFCPTVNSLNLCKNCSITNGSVSPSEQFHKYLPFLNRP 1020  
QY 1021 NIKCPKGLAAVSTSVNLSDQVLA SFMAVHKELKNSODYTEALRAARELAANITADL 1080  
DB 1021 NIKCPKGLAAVSTSVNLSDQVLA SFMAVHKELKNSODYTEALRAARELAANITADL 1080  
QY 1081 RKVPGTDAPEVFPYTTINVFYEQYLLTLPBGLFMLSCLVPTFVASCILGLDRSGLL 1140  
DB 1081 RKVPGTDAPEVFPYTTINVFYEQYLLTLPBGLFMLSCLVPTFVASCILGLDRSGLL 1140  
QY 1141 NLLSTVMLLVDTVGMAIMDSYNVSLINTVSAAGMVEFVSHITRSPALSTKPTWLER 1200  
DB 1141 NLLSTVMLLVDTVGMAIMDSYNVSLINTVSAAGMVEFVSHITRSPALSTKPTWLER 1200  
QY 1201 AKEATISMGSAAVAVANTNLPGLVGLAKAQLQIEFFRLNLTITLGLHGLVFLPV 1260  
DB 1201 AKEATISMGSAAVAVANTNLPGLVGLAKAQLQIEFFRLNLTITLGLHGLVFLPV 1260  
QY 1261 ILSYGPDPVNPALALEQKRAEENVAAMVASCNPHSRVSTADNITYVNSFEGSIKGAGA 1320

Db 1261 ILSYVPDVPNALALOKRAEEAVAAVMVASCNHPHSRVSTADNTIYNHSEGISXAGGA 1320  
 QY 1321 ISNFLPNNGROF 1332  
 Db 1321 ISNFLPNNGROF 1332  
 RESULT 6  
 AAM79168  
 ID AAM79168 standard; protein; 1359 AA.  
 AC AAM79168;  
 DT 06-NOV-2001 (first entry)  
 DE Human protein SEQ ID NO 1830.  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 OS Homo sapiens.  
 PN WO200157190-A2.  
 PD 09-AUG-2001.  
 PF 05-FEB-2001; 2001MO-US004098.  
 XX 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y,  
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW,  
 PI Xue AJ, Yang Y, Wejrtman T, Goodrich R,  
 XX WPI, 2001-476283/51.  
 DR N-PSDB; AAK52301.  
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 PT Claim 20, Page 4213-4216; 6221P; English.  
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78333-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX  
 SQ Sequence 1359 AA;  
 Query Match 99.5%; Score 6872.5; DB 4; Length 1359;  
 Best Local Similarity 97.9%; Pred. NO. 0;  
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY 1 MAEAGLRGWLWMLLRLAQSEPYTTIHQPGYCAFYDECGKNEBLSGSLMTLSNVSCLSN 60  
 Db 1 MAEAGLRGWLWMLLRLAQSEPYTTIHQPGYCAFYDECGKNEBLSGSLMTLSNVSCLSN 60  
 QY 61 TPAKRTTSGHLLILOKICPRLTYGPTOACCSAKOLVSLBASITKALLTRCPACSDNF 120  
 Db 61 TPAKRTTSGHLLILOKICPRLTYGPTOACCSAKOLVSLBASITKALLTRCPACSDNF 120  
 QY 121 VNLHCNHTCSPNOSLFINVTRVAQLGAGOLPAVVAEAFYQHSFAEQSYDCSRVVPAA 180  
 Db 121 VNLHCNHTCSPNOSLFINVTRVAQLGAGOLPAVVAEAFYQHSFAEQSYDCSRVVPAA 180  
 QY 121 VNLHCNHTCSPNOSLFINVTRVAQLGAGOLPAVVAEAFYQHSFAEQSYDCSRVVPAA 180  
 Db 121 VNLHCNHTCSPNOSLFINVTRVAQLGAGOLPAVVAEAFYQHSFAEQSYDCSRVVPAA 180  
 QY 181 ATLAAGTCGCVYGSALCNAQRMWLNFGDGTGNGIAPLDITPHLLPQOAVSGSIQPLNEGV 240  
 Db 181 ATLAAGTCGCVYGSALCNAQRMWLNFGDGTGNGIAPLDITPHLLPQOAVSGSIQPLNEGV 240  
 QY 241 ARCNEQGDVATCSQDCAASCAPAIARQALDSTYVLGQMPGSLVLIILCSVFAVNTI 300  
 Db 241 ARCNEQGDVATCSQDCAASCAPAIARQALDSTYVLGQMPGSLVLIILCSVFAVNTI 300  
 QY 301 LTVGFVAPARBDKSKVNDPKGTSLSDKLSFTHTLLGQFQGMGTWVASMPLTITVLVS 360  
 Db 301 LTVGFVAPARBDKSKVNDPKGTSLSDKLSFTHTLLGQFQGMGTWVASMPLTITVLVS 360  
 QY 361 IPVVALAAGLVPTLTDEVELMSAPNSQARSEKAFHDOFGFPFRTNQVILTAPNRSY 420  
 Db 361 IPVVALAAGLVPTLTDEVELMSAPNSQARSEKAFHDOFGFPFRTNQVILTAPNRSY 420  
 QY 421 RYDLSLLGPRNRSGLIDLDLLELELQERLRLQVSWPBAQRNLSLODICTAPLPNDT 480  
 Db 421 RYDLSLLGPRNRSGLIDLDLLELELQERLRLQVSWPBAQRNLSLODICTAPLPNDT 480  
 QY 481 SLVDCCLNSILQYFQNNRTLLLTANQTLMGQTSQVDMQDHPFYCANAPLTFDQGTALAL 540  
 Db 481 SLVDCCLNSILQYFQNNRTLLLTANQTLMGQTSQVDMQDHPFYCANAPLTFDQGTALAL 540  
 QY 541 SCWADYGAVFPPLAIGYKGYKGDYSEAEALIMTFSLNTPAGDPRLAOKLMEAEFLBEM 600  
 Db 541 SCWADYGAVFPPLAIGYKGYKGDYSEAEALIMTFSLNTPAGDPRLAOKLMEAEFLBEM 600  
 QY 601 RAFQRMAGMFOVTPFAERSLEDEINRTTAEDLPIFATSYIVFLYISLALGSSWSRV 660  
 Db 601 RAFQRMAGMFOVTPFAERSLEDEINRTTAEDLPIFATSYIVFLYISLALGSSWSRV 660  
 QY 661 MVDKATLIGYGVAVVLGAVMAAMGFFSYLGRSSLVLIQVVFVLVSQADNIFFLVE 720  
 Db 661 MVDKATLIGYGVAVVLGAVMAAMGFFSYLGRSSLVLIQVVFVLVSQADNIFFLVE 720  
 QY 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALICFGLALTPPAVRTFALTSGLA 780  
 Db 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALICFGLALTPPAVRTFALTSGLA 780  
 QY 781 IIDFLIOMSAFVALSLDSKROEASRLDVCCCKPELPPPGGSEGLLGFPOKAAAPFL 840  
 Db 781 IIDFLIOMSAFVALSLDSKROEASRLDVCCCKPELPPPGGSEGLLGFPOKAAAPFL 840  
 QY 841 LHWITGVVLLFLALFGVSLVSMCHISVGLQDELALPDOSYLDVFLFLNRFEEVAPV 900  
 Db 841 LHWITGVVLLFLALFGVSLVSMCHISVGLQDELALPDOSYLDVFLFLNRFEEVAPV 900  
 QY 901 YFVTTTGVYFSSBAGNNAICSSAGCNPFSTQKIQYATFPPEQSYLAIPASSVWDDPFDW 960  
 Db 901 YFVTTTGVYFSSBAGNNAICSSAGCNPFSTQKIQYATFPPEQSYLAIPASSVWDDPFDW 960  
 QY 961 LTPSSCCRLYISGPNKDCPSTVNSLNCIKNCMSITMGSVBPVSQFHXYLPWFPLNDP 1020  
 Db 961 LTPSSCCRLYISGPNKDCPSTVNSLNCIKNCMSITMGSVBPVSQFHXYLPWFPLNDP 1020  
 QY 1021 NIKCPKGLAAYSTSVNLTSQGYL-----ASRFMAYH 1053  
 Db 1021 NIKCPKGLAAYSTSVNLTSQGYL-----ASRFMAYH 1053

QY 1054 KPLKKSQDYTEALRAARELAANITADLRKVPCTDPAFEVFPYITTVFPEQYLTITPEGL 1113  
 |||||  
 DB 1081 KPLKKSQDYTEALRAARELAANITADLRKVPCTDPAFEVFPYITTVFPEQYLTITPEGL 1140  
 |||||  
 QY 1114 FMLSCLVFTPAVSCGLIGLDLRSGLLNLTSMILVDVVGFMALMDISYNAVSLINLVS 1173  
 |||||  
 DB 1141 FMLSCLVFTPAVSCGLIGLDLRSGLLNLTSMILVDVVGFMALMDISYNAVSLINLVS 1200  
 |||||  
 QY 1174 AVGMSVEFVSHITRSPFAISTKPTWLERAKEBATISMGSAFVAGVAMTNLPGLIVLGLAKAQ 1233  
 |||||  
 DB 1201 AVGMSVEFVSHITRSPFAISTKPTWLERAKEBATISMGSAFVAGVAMTNLPGLIVLGLAKAQ 1260  
 |||||  
 QY 1234 LQIQIFFRNLNLTITLGLHGLVFLPYIISTYGPDVNPLALBQKAEBAVAAVWVASC 1293  
 |||||  
 DB 1261 LQIQIFFRNLNLTITLGLHGLVFLPYIISTYGPDVNPLALBQKAEBAVAAVWVASC 1320  
 |||||  
 QY 1294 NHPSRVSTADNTYVNHSPFSGSIKAGAGISNPLPNNGROF 1332  
 |||||  
 DB 1321 NHPSRVSTADNTYVNHSPFSGSIKAGAGISNPLPNNGROF 1359  
 |||||  
 RESULT 7  
 ADJ27337  
 ID ADJ27337 standard; protein; 1359 AA.  
 XX  
 AC ADJ27337;  
 XX  
 DT 20-MAY-2004 (first entry)  
 XX  
 DE Human NPC1L1.  
 XX  
 KW Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;  
 KW plasma membrane; transport signal; promoter;  
 KW sterol regulated element binding protein 1; SREBP1;  
 KW binding consensus sequence; transmembrane domain; sterol-sensing domain;  
 KW SSD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;  
 KW cholesterol absorption; serum cholesterol; hyperlipidaemia;  
 KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO2004009772-A2.  
 XX  
 PD 29-JAN-2004.  
 XX  
 PF 17-JUL-2003; 2003WO-US022467.  
 XX  
 PR 19-JUL-2002; 2002US-0397442P.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;  
 XX  
 DR MPI: 2004-132945/13.  
 DR N-PSDB; ADJ27336.  
 DR GENBANK; AF192522.  
 XX  
 PT New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting  
 PT or identifying antagonists of NPC1L1 for inhibiting intestinal  
 PT cholesterol absorption in a subject, or for treating elevated serum  
 PT cholesterol or stroke.  
 XX  
 PS Example 7; SEQ ID NO 44; 125bp; English.  
 XX  
 CC This sequence represents a Niemann-Pick disease, type C1, gene-like 1  
 CC (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains  
 CC a motif ADJ27331 which acts as a trans-golgi network to plasma membrane  
 CC transport signal, and which exhibits limited tissue distribution and  
 CC sterol-mediated abundance. The human NPC1L1 promoter sequence contains a  
 CC sterol regulated element binding protein 1 (SREBP1) binding consensus  
 CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-  
 CC sensing domain (SSD) which is involved in sensing cholesterol levels,  
 CC possibly by a mechanism which involves direct cholesterol bonding. NPC1L1

CC has 42% amino acid sequence homology to human NPC1, a receptor  
 CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or  
 CC polynucleotides are useful for detecting or identifying antagonists of  
 CC NPC1L1, which can be used for inhibiting intestinal cholesterol  
 CC absorption in a subject, or for treating medical conditions including  
 CC elevated serum cholesterol, hyperlipidaemia, atherosclerosis, coronary  
 CC heart disease, stroke or arteriosclerosis.  
 XX  
 SQ Sequence 1359 AA;  
 Query Match 99.5%; Score 6872.5; DB 8; Length 1359;  
 Best Local Similarity 97.9%; Pred. No. 0;  
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;  
 QY 1 MAEAGIRGMLWALLRLAQSSEPYTTHQPGYCAFYDCGKXNPLSGSLMITSNVSCLSN 60  
 DB 1 MAEAGIRGMLWALLRLAQSSEPYTTHQPGYCAFYDCGKXNPLSGSLMITSNVSCLSN 60  
 QY 61 TPARKITGDHLLLOKICPRLYTGPNTQACSAKQVLSBASLITKALTTCPCASDNF 120  
 DB 61 TPARKITGDHLLLOKICPRLYTGPNTQACSAKQVLSBASLITKALTTCPCASDNF 120  
 QY 121 VNLHCNHTCSFNSQSLFINVTRVAQAGAGQLPAVVAVEAFYQHSFAEQSYDSCSRVRVPA 180  
 DB 121 VNLHCNHTCSFNSQSLFINVTRVAQAGAGQLPAVVAVEAFYQHSFAEQSYDSCSRVRVPA 180  
 QY 181 ATLAAGTMCVGYGSAALCNAQRWLNFGDGTGNGLABLDITPHLLEBQAVAGSGIQPLNEGV 240  
 DB 181 ATLAAGTMCVGYGSAALCNAQRWLNFGDGTGNGLABLDITPHLLEBQAVAGSGIQPLNEGV 240  
 QY 241 AHCNSQGDVATCGCCDCAASCRAIARPOLADSTFYIGOMGSLVLTILICSVAVVTI 300  
 DB 241 AHCNSQGDVATCGCCDCAASCRAIARPOLADSTFYIGOMGSLVLTILICSVAVVTI 300  
 QY 301 LTVGRVAPARPKSKMWPBKGTSLSDKLSFSTHTILGQFQSGMTWVASWPLTILVLSV 360  
 DB 301 LTVGRVAPARPKSKMWPBKGTSLSDKLSFSTHTILGQFQSGMTWVASWPLTILVLSV 360  
 QY 361 IPVVALAAGLVFTLITDPVELMSAPNSQAREEKAFHQHPFPFRITNVITABNRSSY 420  
 DB 361 IPVVALAAGLVFTLITDPVELMSAPNSQAREEKAFHQHPFPFRITNVITABNRSSY 420  
 QY 421 RYDSILLGPKNFGSLIDDLLELLELOERLRLQVSPBAQRNLSLDICTAPLNPNPT 480  
 DB 421 RYDSILLGPKNFGSLIDDLLELLELOERLRLQVSPBAQRNLSLDICTAPLNPNPT 480  
 QY 481 SYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANNAPLTRKQGTALNL 540  
 DB 481 SYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANNAPLTRKQGTALNL 540  
 QY 541 SCMDYGAVPFPFLAIGYKGDYSEAEALINTFSLANNYPADPRLAQAKLMEBAFLEBM 600  
 DB 541 SCMDYGAVPFPFLAIGYKGDYSEAEALINTFSLANNYPADPRLAQAKLMEBAFLEBM 600  
 QY 601 RAFORRMAGMFGVTTAERSLDEINRTTAEBLPIFATSYIYIFLYISLALGSSYSWRV 660  
 DB 601 RAFORRMAGMFGVTTAERSLDEINRTTAEBLPIFATSYIYIFLYISLALGSSYSWRV 660  
 QY 661 WYDSKATLGLGVAVALGAVVAAAMGFYSYLGIRSSLVILQVFPVLVSGADNITIFYLE 720  
 DB 661 WYDSKATLGLGVAVALGAVVAAAMGFYSYLGIRSSLVILQVFPVLVSGADNITIFYLE 720  
 QY 721 YQRLRRRPEPRREVIHGRALGRVABSMULCSISEAICFGLALTMPAVRTFALTSGLA 780  
 DB 721 YQRLRRRPEPRREVIHGRALGRVABSMULCSISEAICFGLALTMPAVRTFALTSGLA 780  
 QY 781 IUDFLQMSAFVALLSLDSKROEASRLDYCCVYKQBLPPPOGGGLLIGFQKXAPPL 840  
 DB 781 IUDFLQMSAFVALLSLDSKROEASRLDYCCVYKQBLPPPOGGGLLIGFQKXAPPL 840  
 QY 841 LHMTRGVVTLILFLAFGVSLSYMSCHISVGLDQELALPDSDYLLYPLFLNRYFEVGA 900  
 DB 841 LHMTRGVVTLILFLAFGVSLSYMSCHISVGLDQELALPDSDYLLYPLFLNRYFEVGA 900



D	b	481	SLYDCINSILQYFQNNRLTLLLTANQTLMGQISQYDMKDHFLYCANAPLTFDGTALAL	540
Q	y	541	SCMADYGAVPFPPLAIGGYKQDYSAEALIMTFSLNNTPAGDPRLAQAKLMEAPLEEM	6000
D	b	541	SCMADYGAVPFPFLAIGGYKQDYSAEALIMTFSLNNTPAGDPRLAQAKLMEAPLEEM	6000
Q	y	601	RAFORRMGMFOVTFPAERSLEDEINRTAEDLPFATSYIVFLVYSLALGSYSMSRV	6666
D	b	601	RAFORRMGMFOVTFPAERSLEDEINRTAEDLPFATSYIVFLVYSLALGSYSMSRV	6666
Q	y	661	MVDSKATLGLGAVAVVLGAVMAAMGFYSYIGRSILVIQVVEFLVSVGADNIPLFVLE	7200
D	b	661	MVDSKATLGLGAVAVVLGAVMAAMGFYSYIGRSILVIQVVEFLVSVGADNIPLFVLE	7200
Q	y	721	YQRLPRRPGEPREVIHIGRALGRVAPSMILCSLSEACFELGALTMPAVTFTLSTGLAV	7800
D	b	721	YQRLPRRPGEPREVIHIGRALGRVAPSMILCSLSEACFELGALTMPAVTFTLSTGLAV	7800
Q	y	781	ILDFLLQMSAFALLSLDSKROEASRLDYCCCKPQELPPPGQEGLLGFPQKAAPEL	8400
D	b	781	ILDFLLQMSAFALLSLDSKROEASRLDYCCCKPQELPPPGQEGLLGFPQKAAPEL	8400
Q	y	841	LHMTRGVLLFLFLAFGVSLYSMCHISVGLDDELAPKOSYLLDYFLFNTRYFEGAPV	9000
D	b	841	LHMTRGVLLFLFLAFGVSLYSMCHISVGLDDELAPKOSYLLDYFLFNTRYFEGAPV	9000
Q	y	901	YFVTTILGYNFSSEAGNMNAICSSAGCNNFSTQIKQYATEPEQOSYLAIPASSWVDEIDW	9666
D	b	901	YFVTTILGYNFSSEAGNMNAICSSAGCNNFSTQIKQYATEPEQOSYLAIPASSWVDEIDW	9666
Q	y	961	LTPSSCCRLYITGPNKDFCPSYVNSLNCJKNCMSITMGSYRSVQDFHXYLFWFLNDRP	10200
D	b	961	LTPSSCCRLYITGPNKDFCPSYVNSLNCJKNCMSITMGSYRSVQDFHXYLFWFLNDRP	10200
Q	y	1021	NIKCPKGGLLAAYSTSVNLTSDDQVL-----ASRFAYH	10500
D	b	1021	NIKCPKGGLLAAYSTSVNLTSDDQVL-----ASRFAYH	10500
Q	y	1054	KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVPPYITTNVFEQYILTIPEGL	11100
D	b	1081	KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVPPYITTNVFEQYILTIPEGL	11400
Q	y	1114	FMLSLCLVPTFAVSCLLGLDRLSGLLNLISYMIILVDYVGFMAAMDISTNANSLNLVS	11700
D	b	1141	FMLSLCLVPTFAVSCLLGLDRLSGLLNLISYMIILVDYVGFMAALGISTNANSLNLVS	12000
Q	y	1174	AVGMSVEFASHITRSFAISTKPTWLERABEATISMGSAVPAAGVAMTNLPGLIVLGLAKQ	12300
D	b	1201	AVGMSVEFASHITRSFAISTKPTWLERABEATISMGSAVPAAGVAMTNLPGLIVLGLAKQ	12600
Q	y	1234	LIQIEFFRLNLLITLLGLDGLVFLVILSYVGPDPVNPALALBQKAEBAVVAVMVASCP	12900
D	b	1261	LIQIEFFRLNLLITLLGLDGLVFLVILSYVGPDPVNPALALBQKAEBAVVAVMVASCP	13200
Q	y	1294	NHPSRVSTADNTYVNHSEFGSIXGAGALSNFLPNNGROF	1332
D	b	1321	NHPSRVSTADNTYVNHSEFGSIXGAGALSNFLPNNGROF	1359

RESULT	9
AEB93610	
ID	AEB93610 standard; protein; 1359 AA.
XX	
XX	AEB93610;
AC	
XX	
DT	06-OCT-2005 (first entry)
XX	
DE	Human NPC1L1 protein.
XX	
KM	Protein engineering; NPC1L1; antidiabetic; antiarteriosclerotic;
KM	cardiac; vanotrophic; hyperlipidemia; atherosclerosis; stroke;
KM	arteriosclerosis; ds.

XX	Homo sapiens.
OS	WO2005069900-A2.
PN	04-AUG-2005.
PD	14-JAN-2005; 2005WO-US001469.
PF	16-JAN-2004; 2004US-0537341P.
PR	(MERI ) MERCK & CO INC.
PA	Garcia-Calvo M;
PI	WPI; 2005-564070/57.
DR	N-PDSB; AEB933609.
DR	GENBANK; AF192522.
XX	Identifying ligand of NPC1L1 for stimulating the activity of NPC1L1, by
PT	contacting NPC1L1 with detectably labeled substituted 2-azetidinone
PT	glucuronide and a candidate compound and determining if compound binds to
PT	human NPC1L1.
XX	Disclosure; SEQ ID NO 44; 215bp; English.
PS	The invention relates to identifying a ligand of NPC1L1. The method
XX	involves contacting human NPC1L1 with a detectably labeled substituted 2-
CC	azetidinone glucuronide and a candidate compound and determining whether
CC	the candidate compound binds to human NPC1L1. In identifying a ligand of
CC	NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10
CC	nM or lower. The detectably labeled substituted 2-azetidinone glucuronide
CC	is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands
CC	are useful for stimulating or blocking the activity of NPC1L1, and for
CC	treating conditions caused or mediated by NPC1L1. It is useful for
CC	reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart
CC	disease, stroke, or arteriosclerosis. The present sequence represents the
CC	amino acid sequence of human NPC1L1 as disclosed under Genbank accession
CC	number AF192522.
XX	
XX	
SEQ	Sequence 1359 AA:
Query Match	99.5%; Score 6872.5; DB 9; Length 1359;
Best Local Similarity	97.9%; Pred. No. 0;
Matches 1330; Conservative	0; Mismatches 2; Indels 27; Gaps 1
QY	1 MAEAGLRMTMLALLRLAQSPPYTHIQPGVCAFYDEGKNPEISGLMTLSNNVCSLN 60
DB	1 MAEALRWMWLALLRLQAQSPYTHIQPGVCARYDECGKNPEISGLMTLSNNVCSLN 60
QY	61 TPARKITGDHLILLOKICPLRYTGNTQACSAKOLVSLHNASISTTKALLTRCPACSDNF 120
DB	61 TPARKITGDHLILOKICPLRYLTGNTQACSAKOLVSLHNASISTTKALLTRCPACSDNF 120
QY	121 VNLGHCHTCSPNQSLFIVNTRVAQLGAGQLPAVVAYEAFFYQHSPAEQSYDSCSRVRVPAA 180
DB	121 VNLGHCHTCSPNQSLFIVNTRVAQLGAGQLPAVVAYEAFFYQHSPAEQSYDSCSRVRVPAA 180
QY	181 ATLAVNGTCGYGGSLCNARQLRNQGDTGNKLADLTFLHLBEQANGSGIGPLNEGV 240
DB	181 ATLAAGTCGYGGSLCNARQLRNQGDTGNGLAPLDITFLHLBEQANGSGIGPLNEGV 240
QY	241 ARCNSSOGDDVATCSCODPCAASCPAIAPQALDSFFYLGOMGSLVLIIILCSVFPAWT 300
DB	241 ARCNSSOGDDVATCSCODPCAASCPAIAPQALDSFFYLGOMGSLVLIIILCSVFPAWT 300
QY	301 LLVGRVRAPADSKSMVDPKGTSLSDKLSFSSTHTLLGQFFQMGCTVWASWPJLIIVLSV 360
DB	301 LLVGRVRAPADSKSMVDPKGTSLSDKLSFSSTHTLLGQFFQMGCTVWASWPJLIIVLSV 360
QY	361 IPVVALAAGLVFTETLTPVEIEMSAFNQAREKAFFHOHFGBPFPTNOVILTAPRRSY 420
DB	361 IPVVALAAGLVFTETLTPVEIEMSAFNQAREKAFFHOHFGBPFPTNOVILTAPRRSY 420



```

QY 421 RYDSLLGPKNFSGLLDLDLLELLELOERLRHLQVMSBEAQNISLQDI CYAPLNPDMT 480
  |||
Db 421 RYDSLLGPKNFSGLLDLDLLELLELOERLRHLQVMSBEAQNISLQDI CYAPLNPDMT 480
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTMGQTSQVDMKDHFLYCANAPLTFQGTALAL 540
  |||
Db 481 SLYDCCINSLLQYFQNNRTLLLTANQTMGQTSQVDMKDHFLYCANAPLTFQGTALAL 540
QY 541 SCMDVGAIVPEPLAIGYKKGQYSEAPALIMFESLNNTYPAQBPRLAQAQMLEAPLEEM 600
  |||
Db 541 SCMDVGAIVPEPLAIGYKKGQYSEAPALIMFESLNNTYPAQBPRLAQAQMLEAPLEEM 600
QY 601 RAFORAMAGMFOVTFPAERSLSEDEINRTTAEDLPITATSIVIFLYISLALGSSWSMSRY 660
  |||
Db 601 RAFORAMAGMFOVTFPAERSLSEDEINRTTAEDLPITATSIVIFLYISLALGSSWSMSRY 660
QY 661 MDSKATTLGIGVAVVLGAVMAMGFFSYLGRSSVLIIQVVPFLVLSVGADNIFVLE 720
  |||
Db 661 MDSKATTLGIGVAVVLGAVMAMGFFSYLGRSSVLIIQVVPFLVLSVGADNIFVLE 720
QY 721 YQRLPRRGEPRPVHIGRALGRAPBSMLCSLSEALCFPLGALTMPAVRTFALTGLAV 780
  |||
Db 721 YQRLPRRGEPRPVHIGRALGRAPBSMLCSLSEALCFPLGALTMPAVRTFALTGLAV 780
QY 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPQGGEGLLGFFQKAYAPFL 840
  |||
Db 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPQGGEGLLGFFQKAYAPFL 840
QY 841 LHMITRGVALLFLFALFGVLSYMGCHISVGLDQELALPKDSYLLDYFLPLNRYFEVGAIV 900
  |||
Db 841 LHMITRGVALLFLFALFGVLSYMGCHISVGLDQELALPKDSYLLDYFLPLNRYFEVGAIV 900
QY 901 YPFTTIGVNPSSSAGMNAICSSSAGCNPSFTQKIQVATERPEQSYLAIPASSVNDPEIDM 960
  |||
Db 901 YPFTTIGVNPSSSAGMNAICSSSAGCNPSFTQKIQVATERPEQSYLAIPASSVNDPEIDM 960
QY 961 LFPSSCCRLYISGPNQDKFCPSVTVNSLNCJKNCSITMGSVRSVEQFHXYLPWFLNDRP 1020
  |||
Db 961 LFPSSCCRLYISGPNQDKFCPSVTVNSLNCJKNCSITMGSVRSVEQFHXYLPWFLNDRP 1020
QY 1021 NIKCPKGLAAYSTSVNLTSQGVLT-----ASRFMAYH 1053
  |||
Db 1021 NIKCPKGLAAYSTSVNLTSQGVLT-----ASRFMAYH 1053
QY 1054 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVPEPYITTVFPEQYLTIIPEGL 1113
  |||
Db 1054 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVPEPYITTVFPEQYLTIIPEGL 1113
QY 1081 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVPEPYITTVFPEQYLTIIPEGL 1140
  |||
Db 1081 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVPEPYITTVFPEQYLTIIPEGL 1140
QY 1114 FMLSCLVPTFAVSCLLGLDLSRGLNLTSLYMIIVDTGFPALMDISYNAVSLNLS 1173
  |||
Db 1114 FMLSCLVPTFAVSCLLGLDLSRGLNLTSLYMIIVDTGFPALMDISYNAVSLNLS 1173
QY 1141 FMLSCLVPTFAVSCLLGLDLSRGLNLTSLYMIIVDTGFPALMDISYNAVSLNLS 1200
  |||
Db 1141 FMLSCLVPTFAVSCLLGLDLSRGLNLTSLYMIIVDTGFPALMDISYNAVSLNLS 1200
QY 1174 AVGMSVEPVSHITRSPALISTKPTMLERAKKATISMSAVPAGVAMTNLPGLIYLGLAKQ 1233
  |||
Db 1174 AVGMSVEPVSHITRSPALISTKPTMLERAKKATISMSAVPAGVAMTNLPGLIYLGLAKQ 1233
QY 1201 AVGMSVEPVSHITRSPALISTKPTMLERAKKATISMSAVPAGVAMTNLPGLIYLGLAKQ 1260
  |||
Db 1201 AVGMSVEPVSHITRSPALISTKPTMLERAKKATISMSAVPAGVAMTNLPGLIYLGLAKQ 1260
QY 1234 LIOIPEFRNLMLTTLLGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVVAAMVASC 1293
  |||
Db 1234 LIOIPEFRNLMLTTLLGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVVAAMVASC 1293
QY 1261 LIOIPEFRNLMLTTLLGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVVAAMVASC 1320
  |||
Db 1261 LIOIPEFRNLMLTTLLGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVVAAMVASC 1320
QY 1294 NHPRSVSTADNIYVNSPEGSIKGAGLISNPLFNNGRQF 1332
  |||
Db 1294 NHPRSVSTADNIYVNSPEGSIKGAGLISNPLFNNGRQF 1332
QY 1321 NHPRSVSTADNIYVNSPEGSIKGAGLISNPLFNNGRQF 1359
  |||
Db 1321 NHPRSVSTADNIYVNSPEGSIKGAGLISNPLFNNGRQF 1359

```

RESULT 10  
 ABG22693  
 ID ABG22693 standard; protein; 1344 AA.  
 AC  
 AC ABG22693;  
 XX  
 DT 18-FEB-2002 (first entry)

```

XX XX Novel human diagnostic protein #22684.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS MO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Dmanac RT, Liu C, Tang YT;
XX PI
XX DR N-PSDB; AAS86880.
XX DR WPI: 2001-639362/73.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostic, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 53052; 103pp; English.
XX PS
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probe, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostic as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for creating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ
XX Sequence 1344 AA;
XX
QY Query Match 94.6%; Score 6536; DB 4; Length 1344;
  |||
Db Best local similarity 96.7%; Pred. No. 0;
  |||
QY Matches 12/1; Conservative 7; Mismatches 21; Indels 16; Gaps 4;
  |||
Db
  |||
QY 1 MBEAGLKGWLLMALLRLAQSSEPYTTIHQPGYCAFYDECGKPELSGSLMTLSNVSCLSN 60
  |||
Db 1 MBEAGLKGWLLMALLRLAQSSEPYTTIHQPGYCAFYDECGKPELSGSLMTLSNVSCLSN 60
QY 61 TPARKITGDHLILLOKICPRLYTGPTQACCSAKOLVLSLSTLTKALLTCPCASDNF 120
  |||
Db 61 TPARKITGDHLILLOKICPRLYTGPTQACCSAKOLVLSLSTLTKALLTCPCASDNF 120
QY 121 VNLHGANTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFABQSYDSCSRVRVPAA 180
  |||
Db 121 VNLHGANTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFABQSYDSCSRVRVPAA 180
QY 181 ATLAVGTMGVYGSALCNAQRWLNFGDTGNGLAPLDITFHLLPEQAVSGSIOPLNBEV 240
  |||
Db 181 ATLAVGTMGVYGSALCNAQRWLNFGDTGNGLAPLDITFHLLPEQAVSGSIOPLNBEV 240

```



Db 181 ATLAVTMCVYGSALCNAQRWLNQDGTGNGLAPLDITFHLLEPGQAVSGSIOPLNBSV 240  
QY 241 ARCNESQGDVATCSGDCDCAASC- -A1ARPOLDSTFYLGMPGSLVLIILICGVFAV 298  
Db 241 ARCNESQGDVATCSGDCDCAASCPRPAPRSTPSTW- -RQMPGSLVLIILICGVFAV 298  
QY 299 TLLVGFVAPARADSKVDPKKGTSLSDKLSFSTHTLLGQFQGWVNASMPRLTIVL 358  
Db 299 TLLVGFVAPARADSKVDPKKGTSLSDKLSFSTHTLLGQFQGWVNASMPRLTIVL 358  
QY 359 SVTPVVALAAGLVFELTTPDPELMSAPVSOARSEKAFHDHGPFRFNNVYLTAAPNS 418  
Db 359 SVTPVVALAAGLVFELTTPDPELMSAPVSOARSEKAFHDHGPFRFNNVYLTAAPNS 418  
QY 419 SVRYDSSLGPPKNSFGSLDLDLLELLEQLERLRLQVMSPEAQRNLSQDIQVAPLND 478  
Db 419 SVRYDSSLGPPKNSFGSLDLDLLELLEQLERLRLQVMSPEAQRNLSQDIQVAPLND 478  
QY 479 NTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDTNL 538  
Db 479 NTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDTNL 538  
QY 539 ALSCMADYGAPEPFLAIGYKGYKDYSEAEALIMTFSLNNYPADPRLAQAALMEBAFLE 598  
Db 539 ALSCMADYGAPEPFLAIGYKGYKDYSEAEALIMTFSLNNYPADPRLAQAALMEBAFLE 598  
QY 599 EMRAFORRMAGMFQYTFTAERSLEDEINRTAEDLP1FATS1YIVFLYISLALGSYSWS 658  
Db 599 EMRAFORRMAGMFQYTFTAERSLEDEINRTAEDLP1FATS1YIVFLYISLALGSYSWS 658  
QY 659 RVMVDSKATLGLGVAVVIQAVMAAMGFFSYGIRSSVILQVVPFLVSVGADNIFTV 718  
Db 659 RVMVDSKATLGLGVAVVIQAVMAAMGFFSYGIRSSVILQVVPFLVSVGADNIFTV 718  
QY 719 LEYORLPRRPGPREVHIGRALGRVAPSWMLCSLSAICEFELGALTPMAVAVTPALTSGL 778  
Db 719 LEYORLPRRPGPREVHIGRALGRVAPSWMLCSLSAICEFELGALTPMAVAVTPALTSGL 778  
QY 779 AVILDFLLQMSAFVALLSDSKROEASRLDVCVCVPOELPPPGQEGELGFFQKAVAP 838  
Db 779 AVILDFLLQMSAFVALLSDSKROEASRLDVCVCVPOELPPPGQEGELGFFQKAVAP 838  
QY 839 FLIHHITRGV- - -LLFLALPGVSLVSMCHI SVGLDOELAPKDSYLLDYFLPLNRYE 894  
Db 839 FLIHHITRGV- - -LLFLALPGVSLVSMCHI SVGLDOELAPKDSYLLDYFLPLNRYE 894  
QY 895 EVGAIVYFTTTLGNPSSBAGNNALCSSAGCNNSFTOKIOVATEPPEOSYLAIPASSMV 954  
Db 895 EVGAIVYFTTTLGNPSSBAGNNALCSSAGCNNSFTOKIOVATEPPEOSYLAIPASSMV 954  
QY 955 DDFIDMLTPSSCCRLYISGPNKDKCPSTVNSLNCIKNCKMSITMGSVRPSVEQFHXYLPW 1014  
Db 955 DDFIDMLTPSSCCRLYISGPNKDKCPSTVNSLNCIKNCKMSITMGSVRPSVEQFHXYLPW 1014  
QY 1015 FLNDRPNIKCPKGLAANSTVNLSDQVTLASREMAHKEPKNSQDYTEALRAABELAA 1074  
Db 1015 FLNDRPNIKCPKGLAANSTVNLSDQVTLASREMAHKEPKNSQDYTEALRAABELAA 1074  
QY 1075 NITATLRKVPGDPAFEVEPPYITVNFVBOYLTLLPEGLFMLSCLVPRFAVSCILLGD 1134  
Db 1075 NITATLRKVPGDPAFEVEPPYITVNFVBOYLTLLPEGLFMLSCLVPRFAVSCILLGD 1134  
QY 1135 LRSGLNLNLSTVMIIVDTVGFMAIWDISYNAVSLINLVSAGMSVEFVSHIRSFALSTK 1194  
Db 1135 LRSGLNLNLSTVMIIVDTVGFMAIWDISYNAVSLINLVSAGMSVEFVSHIRSFALSTK 1194  
QY 1195 PTMLERAKAATISMSGSAVAGVAMTNVPGIIVLGLAKAQLIOI FFFRLNLLITLLGLLHG 1254  
Db 1195 PTMLERAKAATISMSGSAVAGVAMTNVPGIIVLGLAKAQLIOI FFFRLNLLITLLGLLHG 1254  
QY 1255 LVFLVLIISYVGPDPNPALALEOKABEEA- - - - -VAAVNVASGPNRPSVST 1301  
Db 1255 LVFLVLIISYVGPDPNPALALEOKABEEA- - - - -VAAVNVASGPNRPSVST 1301  
QY 1301 LVFLVLIISYVGPDPNPALALEOKABEEA- - - - -VAAVNVASGPNRPSVST 1301  
Db 1301 LVFLVLIISYVGPDPNPALALEOKABEEA- - - - -VAAVNVASGPNRPSVST 1301

RESULT 11  
ADJ27295  
ID ADJ27295 standard; proteiin; 1331 AA.  
XX  
AC ADJ27295;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Rat NPC1L1.  
XX  
KW Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;  
KW plasma membrane; transport signal; promoter;  
KW sterol regulated element binding protein 1; SREBP1;  
KW binding consensus sequence; transmembrane domain; sterol-sensing domain;  
KW SSD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;  
KW cholesterol absorption; serum cholesterol; hyperlipidaemia;  
KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.  
XX  
OS Rattus sp.  
XX  
PN MO200409772-A2.  
XX  
PD 29-JAN-2004.  
XX  
PF 17-JUL-2003; 2003WO-US022467.  
XX  
PR 19-JUL-2002; 2002US-0397442P.  
XX  
PA (SCHE ) SCHERING CORP.  
XX  
PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;  
XX  
DR MPI: 2004-132945/13.  
XX  
DR N-PSDB: ADJ27294.  
XX  
PT New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting  
PT or identifying antagonists of NPC1L1 for inhibiting intestinal  
PT cholesterol absorption in a subject, or for treating elevated serum  
PT cholesterol or stroke.  
XX  
PS Claim 1; SEQ ID NO 2; 125bp; English.  
XX  
XX  
CC This sequence represents a Niemann-Pick disease, type C1, gene-like 1  
CC (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains  
CC a motif ADJ27331 which acts as a trans-golgi network to plasma membrane  
CC transport signal, and which exhibits limited tissue distribution and  
CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a  
CC sterol regulated element binding protein 1 (SREBP1) binding consensus  
CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-  
CC sensing domain (SSD) which is involved in sensing cholesterol levels,  
CC possibly by a mechanism which involves direct cholesterol binding. NPC1L1  
CC has 428 amino acid sequence homology to human NPC1, a receptor  
CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or  
CC polynucleotides are useful for detecting or identifying antagonists of  
CC NPC1L1, which can be used for inhibiting intestinal cholesterol  
CC absorption in a subject, or for treating medical conditions including  
CC elevated serum cholesterol, hyperlipidaemia, arteriosclerosis, coronary  
CC heart disease, stroke or arteriosclerosis.  
XX  
SQ Sequence 1331 AA;  
XX  
Query Match 78.5%; Score 5421.5; DB 8; Length 1331;  
Beet Local Similarity 77.8%; Pred. No. 0;  
Matches 1039; Conservative 120; Mismatches 165; Indels 7; Gaps 6;  
QY 1 MAEAGLRGMLMALRLAAGEPYTHQPCYCAFYDECGNPKLSGLMTLSNVSCLSN 60  
Db 1 MAEAGLRGMLMALRLAAGEPYTHQPCYCAFYDECGNPKLSGLMTLSNVSCLSN 59  
QY 61 TPARKITGDHLLLOKTRPRITGPNTQ-ACCSAKQVLSLASLSTTALLTRCAGCDN 119  
Db 61 TPARKITGDHLLLOKTRPRITGPNTQ-ACCSAKQVLSLASLSTTALLTRCAGCDN 119

Db 60 TEARHVTGHEHLALLORI CPELYNGPNTTACCSYKQLLSLESSMSITTKALLTFCPCASDN 119  
 Qy 120 FYNLHCHNTCSPNQSLFINVTRVAOLGAGQLPAVVAVEAFYCHSPAESYDSCSRVRPA 179  
 Db 120 FVSLHCHNTCSPNQSLFINVTRVERKAGEPPAVVAVEAFYCHSPAEKATBSCSQRIRPA 179  
 Qy 180 AATLAVGTMGVYGSALCNANORMLNFQDGTNGIAPLDITPHLLEPQAVGSGIQPLMRG 239  
 Db 180 AASLAVGSMGVYGSALCNANORMLNFQDGTNGIAPLDITPHLLEPQALPDQIQPLNGK 239  
 Qy 240 VARCNESQGDVATCCODCAASCRAIARQALDSTFYLGOMGSLVLIILCSVAVVT 299  
 Db 240 IAPCNESQGDSDSAYVCSODCAASCRAIARPEBALRPFYMGKRMGMWALIIIFPAVFLVS 299  
 Qy 300 ILLVGRVAPARDKSMVDPKKGTSLSDSLSTFHTLLGQFQGMGTWVASWPLTIVLS 359  
 Db 300 AVLVRLRVVSNRKNKAABQAPQAPKLPKHKLSPHIIIGRFFQNMGTWVASWPLTIVLALS 359  
 Qy 360 VLPVVALAAGLVETELTTPVELMSAPNSQARSEKAFFHQHFGPFRTNOVILTAPNRS 419  
 Db 360 FIVVIALAAGLVETELTTPVELMSAPKSGARKEKSFHDEHFGPFRTNOIIPVTAANRS 419  
 Qy 420 YKYSLSLLPKNFSGLLDDLLLELQERLNLQVSWPFAORNI SLQDTCYAPLNP 479  
 Db 420 YKYSLSLLGSKNFSGLSLDFLELELQERLNLQVSWPFAORNI SLQDTCYAPLNP 479  
 Qy 480 TSLYDCINSLQYFQNNRTLLLTANQTMGOTQVDMQDHLKYCANAPLTFKDGTA 539  
 Db 480 TSLSDCVNSLQYFQNNRTLLMLTANQTMGOTSLVDMQDHLKYCANAPLTFKDGTA 539  
 Qy 540 LSCMDYGAVPPEFLAIGYKGYSEAEALIMTFSLNNYPAGDPRLAQKMEAEFLER 599  
 Db 540 LSCMDYGAVPPEFLAVGQYQGYSEAEALITFSLNNYPADDPMAQKMEAEFLER 599  
 Qy 600 MEAFQRMAGMFOVTTAERSLDEINRTTADLPPIATSYIVLYISIALGYSYSWSR 659  
 Db 600 MESFGNTSDKFOVAFSAERSLDEINRTTQDLPVFAVSYIIVLYISIALGYSYRCSR 659  
 Qy 660 VAVDSKATGLGIVANVLGAVMAAMGFYSYIGRSSVLVQVPPVLSGAMNITFVL 719  
 Db 660 VAVESKATGLGIVAVLGAVLAAMGFYSYIGRSSVLVQVPPVLAAGAMNITFVL 719  
 Qy 720 EYQRLPRRGEPEPEVAHIGRALGRVAPSMILCSLSEALICFGLGALTMPAVRTALTSGLA 779  
 Db 720 EYQRLPRMGEORAEHIGRTLSVAPSMILCSLSEALICFGLGALTMPAVRTALTSGLA 779  
 Qy 780 VILDFLLQNSAFVALLSLDSKROEASRLDYCCCKVQDELPPQDGEGLLGFQKAYAP 839  
 Db 780 IILDFLLQNTAFVALLSLDSKROEASRPDVLCCFSYTRKLPPEKKEGLLRFPRKIYAP 839  
 Qy 840 LLHMTIRGVVLLFLALFGVSLYSMGHISVGLDQELALPRDSVILDYFLANRYFVGAP 899  
 Db 840 LLHRTIRPVVMLLFLFLGANLYLMCNINVGLDQELALPRDSVILDYFLANRYLEVGP 899  
 Qy 900 VYFVTLTYGVNFSSEAGMNAICSSAGCNFSFTOKIOVATFPPQSYLAIAPASSWVDFID 959  
 Db 900 VYFVTLTYGVNFSSEAGMNAICSSAGCKSBSLTOKIOVASEFPQSYVALAASSWVDFID 959  
 Qy 960 WLP-TSP-SSCRLYISGNKDKFCPSTVNSLNCNKSITMGSVRPSVBOFHXYLPWFLND 1018  
 Db 960 WLPSSSCRLYIRGBHDEFCPSTDTSEFCNKNCRITLGPVRAFAEQHXYLPWFLND 1019  
 Qy 1019 RNRIRKPKGGLAAYSTSVNLTSPOYLASFMAVHPRKNSQOYTERALRAELANITA 1078  
 Db 1020 PPRIRKPKGGLAAYSTSVNLSDGOVIASOFMAVHPRKNSQOFTELARSLANITA 1079  
 Qy 1079 DLKRVGTGPAPFVPYTTTNNVYBOYLTLPEGLFMLSCLVPTFAVSCILGLDLRS 1138  
 Db 1080 DLKRVGTGPBNFVPYTTTNNVYQOYLTLPEGLFMLSCLVPTFAVSCILGLDMCS 1139  
 Qy 1139 LNLLSIWLVDVTFGFMALMDSIYNAVSLINIVSAVGSVEFVSHITRSFAISTXPTWL 1198  
 Db 1140 LNLLSIWLVDVTFGFMALMDSIYNAVSLINIVTAVGSVEFVSHITRSFAISTXPTWL 1199

Qy 1199 ERAKEATISMSAIVFAGVAMTNLPGILVGLAKAQLIQIFPRLNLLITLLGLHGLVPL 1258  
 Db 1200 ERAKDAITVMGSAIVFAGVAMTNFPGILILGFAQAQLIQIFPRLNLLITLLGLHGLVPL 1259  
 Qy 1259 PVTLSTVGPDPVDPALBEOGRABEAVAAVMAVASCNHPBSRVSTADNIYNNHSESGSI-KG 1317  
 Db 1260 PVTLSTVGPDPVDPALVQOEKLASEA-AVAPBSQCPQYPSPADAN--VNYGFAPELARG 1316  
 Qy 1318 AGAISNFPNNROF 1332  
 Db 1317 ANARSLSPKSDQF 1331  
 RESULT 12  
 ADY60738  
 ID ADY60738 standard; protein; 1331 AA.  
 XX  
 AC ADY60738;  
 XX  
 DT 19-MAY-2005 (first entry)  
 XX  
 DE Rat Niemann-Pick C1-like protein 1 (NPC1L1) protein.  
 XX  
 KW Niemann-Pick C1-like protein 1; Antilipemic; Antiarteriosclerotic;  
 KW Cardiant; Vasotropic; Cerebroprotective; cholesterol; hyperlipidemia;  
 KW metabolic disorder; atherosclerosis; cardiovascular disease;  
 KW coronary artery disease; cerebrovascular ischemia; neurological disease;  
 KW arteriosclerosis; transgenic animal.  
 XX  
 OS Rattus sp.  
 XX  
 PN W02005015986-A1.  
 XX  
 PD 24-FEB-2005.  
 XX  
 PF 16-DEC-2003; 2003WC-US040113.  
 XX  
 PR 17-JUL-2003; 2003US-00621758.  
 PR 22-AUG-2003; 2003US-00646301.  
 PR 16-SEP-2003; 2003US-00663208.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;  
 XX  
 DR MPI; 2005-284403/29.  
 XX  
 DR N-PSDB; ADY60737, ADY60746.  
 XX  
 PT Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of  
 PT NPC1L1 that inhibits NPC1L1 mediated sterol or 5 alpha-sterol uptake, for  
 PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.  
 XX  
 PS Claim 30; SEQ ID NO 2; 146p; English.  
 XX  
 CC The invention relates to an isolated Niemann-Pick C1-like protein 1  
 CC (NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an  
 CC amino acid sequence appearing as ADY60738,ADY60740 or ADY60748 (Rat,  
 CC human and Mouse NPC1L1). Also included are an isolated polynucleotide  
 CC encoding NPC1L1 (appearing as ADY60737,ADY60739 or ADY60747), a  
 CC recombinant vector comprising the polynucleotide, a host cell comprising  
 CC the vector, an isolated antibody which specifically binds to NPC1L1 (or  
 CC to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),  
 CC preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous  
 CC mutation of endogenous chromosomal NPC1L1 (where the mouse does not  
 CC produce any functional NPC1L1 protein), an offspring or progeny of the  
 CC mouse having inherited a mutated NPC1L1 allele of the mouse, screening a  
 CC sample for an intestinal sterol or Salpha-sterol absorption antagonist,  
 CC inhibiting NPC1L1 mediated sterol or Salpha-sterol uptake in a subject, a  
 CC kit (comprising aecetamide or substituted azetidione in a pharmaceutical  
 CC dosage form, and information indicating that NPC1L1 is a target of  
 CC aecetamide or substituted azetidione), decreasing the level of intestinal  
 CC sterol or Salpha-sterol absorption in a subject (involving reducing the



PT glucuronide and a candidate compound and determining if compound binds to  
human NPC1L1.

XX Example 2; SEQ ID NO 2; 215bp; English.

CC The invention relates to identifying a ligand of NPC1L1. The method  
CC involves contacting human NPC1L1 with a detectably labeled substituted 2-  
CC azetidine glucuronide and a candidate compound and determining whether  
CC the candidate compound binds to human NPC1L1. In identifying a ligand of  
CC NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10  
CC nM or lower. The detectably labeled substituted 2-azetidine glucuronide  
CC is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands  
CC are useful for stimulating or blocking the activity of NPC1L1, and for  
CC creating conditions caused or mediated by NPC1L1. It is useful for  
CC reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart  
CC disease, stroke, or arteriosclerosis. The present sequence represents a  
CC rat NPC1L1, a N-glycosylated protein.

XX Sequence 1331 AA;

Query Match 78.5%; Score 5421.5; DB 9; Length 1331;  
Best Local Similarity 77.8%; Pred. No. 0;  
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MBEAGLRGWLWALLRLAQSEPTTTHQPGYCAFYDECGKNBELSGIMTISNVCISN 60  
DB 1 MAAAWL-GWLWALLRLSAAGELYTPKHEAGVCTFEECGKNBELSGILTSINVCISN 59  
QY 61 TPAKRTGHLILQKICPRLYTGENTQ-ACCSAKOLVISEASLTSTKALTTRCPACSN 119  
DB 60 TPAAHVTGHLALQKICPRLYNGPNTTACCSKTLISSEMSITKALTTRCPACSN 119  
QY 120 FNLHCHNTCSPNQSFINTVRAOAGOLPAVVAVEAFYQHSFQESYDCSRARVA 179  
DB 120 FNLHCHNTCSPNQSFINTVRAOAGOLPAVVAVEAFYQHSFQESYDCSRARVA 179  
QY 120 FNLHCHNTCSPNQSFINTVRAOAGOLPAVVAVEAFYQHSFQESYDCSRARVA 179  
DB 120 FNLHCHNTCSPNQSFINTVRAOAGOLPAVVAVEAFYQHSFQESYDCSRARVA 179  
QY 180 AATLAVTGCYVYGSALCNARWLNFGDGTNGLABLDITFHLBEGAVSGGIQPLNBG 239  
DB 180 AATLAVTGCYVYGSALCNARWLNFGDGTNGLABLDITFHLBEGAVSGGIQPLNBG 239  
QY 180 AATLAVTGCYVYGSALCNARWLNFGDGTNGLABLDITFHLBEGAVSGGIQPLNBG 239  
DB 180 AATLAVTGCYVYGSALCNARWLNFGDGTNGLABLDITFHLBEGAVSGGIQPLNBG 239  
QY 240 VARNESQGDVATCGCCDCAACCAIARPAOLDSTFYIGOMGSLVLIILCSVAAYT 299  
DB 240 VARNESQGDVATCGCCDCAACCAIARPAOLDSTFYIGOMGSLVLIILCSVAAYT 299  
QY 240 VARNESQGDVATCGCCDCAACCAIARPAOLDSTFYIGOMGSLVLIILCSVAAYT 299  
DB 240 VARNESQGDVATCGCCDCAACCAIARPAOLDSTFYIGOMGSLVLIILCSVAAYT 299  
QY 300 ILLVGRVAPARKSMVNPBKGTSLSDKLSFTHTLLGQFGQWGTWVASWPLTIVLS 359  
DB 300 ILLVGRVAPARKSMVNPBKGTSLSDKLSFTHTLLGQFGQWGTWVASWPLTIVLS 359  
QY 300 ILLVGRVAPARKSMVNPBKGTSLSDKLSFTHTLLGQFGQWGTWVASWPLTIVLS 359  
DB 300 ILLVGRVAPARKSMVNPBKGTSLSDKLSFTHTLLGQFGQWGTWVASWPLTIVLS 359  
QY 360 VLPVVALAAGLVTEHTTDPVELMSAPNSQARSEKAFHDQFGPFRITNOVILTAENRS 419  
DB 360 VLPVVALAAGLVTEHTTDPVELMSAPNSQARSEKAFHDQFGPFRITNOVILTAENRS 419  
QY 360 VLPVVALAAGLVTEHTTDPVELMSAPNSQARSEKAFHDQFGPFRITNOVILTAENRS 419  
DB 360 VLPVVALAAGLVTEHTTDPVELMSAPNSQARSEKAFHDQFGPFRITNOVILTAENRS 419  
QY 420 YRYSLLGPKNFGSLIDLDLLELLELQERLRLHLOVMSPEAORNISLODICYAPLNP 479  
DB 420 YRYSLLGPKNFGSLIDLDLLELLELQERLRLHLOVMSPEAORNISLODICYAPLNP 479  
QY 420 YRYSLLGPKNFGSLIDLDLLELLELQERLRLHLOVMSPEAORNISLODICYAPLNP 479  
DB 420 YRYSLLGPKNFGSLIDLDLLELLELQERLRLHLOVMSPEAORNISLODICYAPLNP 479  
QY 480 TSLYDCISLLOYPQNNRTLLLTNQTLMGOTQVDMQDHLXCANAPLTFKDGSTALA 539  
DB 480 TSLYDCISLLOYPQNNRTLLLTNQTLMGOTQVDMQDHLXCANAPLTFKDGSTALA 539  
QY 480 TSLYDCISLLOYPQNNRTLLLTNQTLMGOTQVDMQDHLXCANAPLTFKDGSTALA 539  
DB 480 TSLYDCISLLOYPQNNRTLLLTNQTLMGOTQVDMQDHLXCANAPLTFKDGSTALA 539  
QY 540 LSCMADYGAVPFPLAIGYKGYSEAEALINTFSLNNYPADGPLAOKLMEAEFLBE 599  
DB 540 LSCMADYGAVPFPLAIGYKGYSEAEALINTFSLNNYPADGPLAOKLMEAEFLBE 599  
QY 540 LSCMADYGAVPFPLAIGYKGYSEAEALINTFSLNNYPADGPLAOKLMEAEFLBE 599  
DB 540 LSCMADYGAVPFPLAIGYKGYSEAEALINTFSLNNYPADGPLAOKLMEAEFLBE 599  
QY 600 MRAFORMAGMFOVTTAERSLEDEINRTAEDLPITANSYIVIFLYISIALGYSYSWSR 659  
DB 600 MRAFORMAGMFOVTTAERSLEDEINRTAEDLPITANSYIVIFLYISIALGYSYSWSR 659  
QY 600 MRAFORMAGMFOVTTAERSLEDEINRTAEDLPITANSYIVIFLYISIALGYSYSWSR 659  
DB 600 MRAFORMAGMFOVTTAERSLEDEINRTAEDLPITANSYIVIFLYISIALGYSYSWSR 659  
QY 660 VMDVSAATLGLGVAVVILGAVVAMAGFESYLGRSSIVILQVAPVILSVGANITFLV 719  
DB 660 VMDVSAATLGLGVAVVILGAVVAMAGFESYLGRSSIVILQVAPVILSVGANITFLV 719  
QY 660 VMDVSAATLGLGVAVVILGAVVAMAGFESYLGRSSIVILQVAPVILSVGANITFLV 719  
DB 660 VMDVSAATLGLGVAVVILGAVVAMAGFESYLGRSSIVILQVAPVILSVGANITFLV 719  
QY 720 EYORLPPEPGEQREAHIGRTIGSVAPSMILCSLSAICFPLGALTMPAVRTFALTSGLA 779  
DB 720 EYORLPPEPGEQREAHIGRTIGSVAPSMILCSLSAICFPLGALTMPAVRTFALTSGLA 779

DB 720 EYORLPPEPGEQREAHIGRTIGSVAPSMILCSLSAICFPLGALTMPAVRTFALTSGLA 779  
QY 780 VILDFLQMSAFVALLSIDSKROEASRLDVCCCVARPOELPRPQEGELIGFQRAYAVP 839  
DB 780 VILDFLQMSAFVALLSIDSKROEASRLDVCCCVARPOELPRPQEGELIGFQRAYAVP 839  
QY 840 LHMWTRGVALLFLALFGVSLYSMCHISVGLDQELAPKDSYLLDYLFLNRYEVCAP 899  
DB 840 LHMWTRGVALLFLALFGVSLYSMCHISVGLDQELAPKDSYLLDYLFLNRYEVCAP 899  
QY 900 VYFVTTLGNFSESGMNAICSSAGCNPSFQKIOYATEPEQSYLAIIPASSWDDPFD 959  
DB 900 VYFVTTLGNFSESGMNAICSSAGCNPSFQKIOYATEPEQSYLAIIPASSWDDPFD 959  
QY 960 WLP-SSCCRLYISGPNKDFCPSYVNSLNCCKMSTMGSVRSVEQFHYLPMFLND 1018  
DB 960 WLP-SSCCRLYISGPNKDFCPSYVNSLNCCKMSTMGSVRSVEQFHYLPMFLND 1018  
QY 960 WLP-SSCCRLYISGPNKDFCPSYVNSLNCCKMSTMGSVRSVEQFHYLPMFLND 1018  
DB 960 WLP-SSCCRLYISGPNKDFCPSYVNSLNCCKMSTMGSVRSVEQFHYLPMFLND 1018  
QY 1019 RPNKCPKGLAAVSTVNLTSQGVLASRFMAVHKPLKNSQDYTEALRADELAAITA 1078  
DB 1019 RPNKCPKGLAAVSTVNLTSQGVLASRFMAVHKPLKNSQDYTEALRADELAAITA 1078  
QY 1020 PNTNCPKGLAAVSTVNLTSQGVLASRFMAVHKPLKNSQDYTEALRADELAAITA 1079  
DB 1020 PNTNCPKGLAAVSTVNLTSQGVLASRFMAVHKPLKNSQDYTEALRADELAAITA 1079  
QY 1079 DLRKTPGTDPAFVPPYTTTNNFYQYLTILBGLPMLSLCVFPFPAVSCILIGDLSG 1138  
DB 1079 DLRKTPGTDPAFVPPYTTTNNFYQYLTILBGLPMLSLCVFPFPAVSCILIGDLSG 1138  
QY 1080 DLRKTPGTDPAFVPPYTTTNNFYQYLTILBGLPMLSLCVFPFPAVSCILIGDLSG 1139  
DB 1080 DLRKTPGTDPAFVPPYTTTNNFYQYLTILBGLPMLSLCVFPFPAVSCILIGDLSG 1139  
QY 1139 LNTLSIWMILVDTGFMALMDISYNAVSLNLYSVAGMSVFEVSHITRSPASTKPTL 1198  
DB 1139 LNTLSIWMILVDTGFMALMDISYNAVSLNLYSVAGMSVFEVSHITRSPASTKPTL 1198  
QY 1140 LNTLSIWMILVDTGFMALMDISYNAVSLNLYSVAGMSVFEVSHITRSPASTKPTL 1199  
DB 1140 LNTLSIWMILVDTGFMALMDISYNAVSLNLYSVAGMSVFEVSHITRSPASTKPTL 1199  
QY 1199 ERAKATISWGSVAVGAVMTNLPGLVYLGLAKAOLQIFFERMLNLTLLGLHGLVFL 1258  
DB 1199 ERAKATISWGSVAVGAVMTNLPGLVYLGLAKAOLQIFFERMLNLTLLGLHGLVFL 1258  
QY 1200 ERAKATISWGSVAVGAVMTNLPGLVYLGLAKAOLQIFFERMLNLTLLGLHGLVFL 1259  
DB 1200 ERAKATISWGSVAVGAVMTNLPGLVYLGLAKAOLQIFFERMLNLTLLGLHGLVFL 1259  
QY 1259 PYLISYVGPVDPALALBQKRAEVAAMVAMASCPNHRSTANTINYNHSEST-KG 1317  
DB 1259 PYLISYVGPVDPALALBQKRAEVAAMVAMASCPNHRSTANTINYNHSEST-KG 1317  
QY 1260 PVLSTYLGPDVQALVOBEKLASEA-AVAPSPSCQYPSPADADAN--VNGFAPBELAHG 1316  
DB 1260 PVLSTYLGPDVQALVOBEKLASEA-AVAPSPSCQYPSPADADAN--VNGFAPBELAHG 1316  
QY 1318 AGAISNPLPNNROF 1332  
DB 1318 AGAISNPLPNNROF 1332  
QY 1317 ANAARSLPKSDQK 1331  
DB 1317 ANAARSLPKSDQK 1331  
RESULT 14  
ADJ27305  
ID ADJ27305 standard; protein; 1333 AA.  
XX AC ADJ27305;  
XX DT 20-MAY-2004 (first entry)  
XX DE Mouse NPC1L1.  
XX NI Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;  
XX PL plasma membrane; transport signal; promoter;  
XX ST sterol regulated element binding protein 1; SREBP1;  
XX BI binding consensus sequence; transmembrane domain; sterol-sensing domain;  
XX SS SCD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;  
XX KW cholesterol absorption; serum cholesterol; hyperlipidemia;  
XX KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.  
OS Mus sp.  
XX WO2004009772-A2.  
XX PD 29-JAN-2004.  
XX PF 17-JUL-2003; 2003WO-US022467.  
XX PR 19-JUL-2002; 2002US-0397442P.  
XX PA (SCHE) SCHERING CORP.

XX Altmann SW, Murgolo NJ, Wang LQ, Graziano MP,  
 XX MPI; 2004-132945/13.  
 DR N-PSDB; ADJ27304.  
 XX  
 PT New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting  
 PT or identifying antagonists of NPC1L1 for inhibiting intestinal  
 PT cholesterol absorption in a subject, or for treating elevated serum  
 PT cholesterol or stroke.  
 XX  
 PS Claim 1, SEQ ID NO 12; 125pp; English.  
 XX  
 CC This sequence represents a Niemann-Pick disease, type C1, gene-like 1  
 CC (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains  
 CC a motif ADJ27331, which acts as a trans-golgi network to plasma membrane  
 CC transport signal, and which exhibits limited tissue distribution and  
 CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a  
 CC sterol regulated element binding protein 1 (SREBP1) binding consensus  
 CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-  
 CC sensing domain (SSD) which is involved in sensing cholesterol levels,  
 CC possibly by a mechanism which involves direct cholesterol bonding. NPC1L1  
 CC has 428 amino acid sequence homology to human NPC1, a receptor  
 CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or  
 CC polynucleotides are useful for detecting or identifying antagonists of  
 CC NPC1L1, which can be used for inhibiting intestinal cholesterol  
 CC absorption in a subject, or for treating medical conditions including  
 CC elevated serum cholesterol, hyperlipidaemia, atherosclerosis, coronary  
 CC heart disease, stroke or arteriosclerosis.  
 CC  
 XX Sequence 1333 AA;  
 XX

Query Match 78.3%; Score 5407; DB 8; Length 1333;  
 Best Local Similarity 77.2%; Pred. No. 0;  
 Matches 1028; Conservative 126; Mismatches 174; Indels 4; Gaps 4;

QY 4 AGRGHMLALLRLAQSPPYTHHOPGYCAFDGCKRPELSSGMTSNISCLSNTPA 63  
 DB 3 AAMOGMLALLRLINSOGELVTPHKAQCTPYEBCGKNPELSSGTLTSSNLSCLSNTPA 62  
 QY 64 RITGDHLLKQICRPLVTGPN-TQACSAKQVLSEASLSTTKALLRCPACSNPNFN 122  
 DB 63 RHVTGDHLLKQVCRPLVNGPDYACCTKQVLSLSSITTKALLRCPACSENPNF 122  
 QY 123 IHCNHTCSPNOSLFINVTVAQLGAGQLPAVVAVEAFYQHSFPAEOSYDSCSRVPAAPAT 182  
 DB 123 IHCNHTCSPDQSLFINVTVAQDQQLPAVVAVEAFYQHSFPAEKAYECSVRIPAPAS 182  
 QY 183 LAVGTCGYGAGALCAQGMWLNFGDGTGNGLAPLDTTHLEBGAQVSGIOPLNEGVAR 242  
 DB 183 LAVGSCGYGAGALCAQGMWLNFGDGTGNGLAPLDTTHLEBGAQVSGIOPLNEGVAR 242  
 QY 243 CNEGSGDVATCSGDCDCAASCPAIAKPAOLDSFVYCGMGSVLIIITCSFPAVVTIL 302  
 DB 243 CNEGSGEDSAACSCDCAASCPVIPPALRPSFTYGRMPGMALIIITTAFFVLSVL 302  
 QY 303 VGFRTVAAPADSKMVDPKKGTSLSDXLSEFTHTLLGQFQGWGTWASWPLTILVLSVP 362  
 DB 303 VTLRTVANSNNKTKKTSQGAAPNLPRKRRTSPHTVLGRFPESNCTRASPLVTLALSFIV 362  
 QY 363 VVLAAGLVFTLTDVPELMSAPNSQARSEKAFHDQHPFFRTNQVILTAPNRSYRY 422  
 DB 363 VIALSVGLTFEILTDPVELMSAPKSQARKEKAFHEHGFPPRTNQIFVTAKNMSYKY 422  
 QY 423 DILLGPKRPSGILDDILLLELELQERLRLQVMSPEAORNTSLDICTYAPLNPNTSL 482  
 DB 423 DILLGPKRPSGILSDLLQELLEQERLRHLQVMSHEQORNTSLDICTYAPLNPNTSL 482  
 QY 483 YDCINSLIQYQNNNTLLLTANQTMQGTQVMDKDFLYCANAPLTFKDGATALASC 542  
 DB 483 TDCCNSSLQYQNNHTLLLTANQTLNGQTSYVMDKDFLYCANAPLTYKQGTALASC 542  
 QY 543 MADYGAPVFPFLAIGSYKGDYSSEALIMTSLNNYPAGDPRLAQAKLMEAFLEEMRA 602

DB 543 IADYGAPVFPFLAIGSYGTDYSEAEALITFSINNYPADDPKMAAKLMEBAFLKEMQS 602  
 QY 603 FORNAGMFOVYFTAEBSLEDEINRTAEDEPIFATSYVIFLYISLALGSSYSSRWVY 662  
 DB 603 FORSTADKFOQLAFASERLEDEINRTTQDLPVFATISLYIFLYISLALGSSYSSRWVAV 662  
 QY 663 DSKATLIGLGAUVVIGAMMAAMGPFSSYIGIRSSLYLDVVPFLVUSVGNINIFIVLEXY 722  
 DB 663 DSKATLIGLGAUVVIGAVVAAAMGFSSYLYGVSSLYLDVVPFLVAVGADNIFIVLEXY 722  
 QY 723 RLPRRPGSPREVIHGRALGRVAPSWLGLSLSEALCFGLATLPMNAVTPALTSGLAYIL 782  
 DB 723 RLPRRPGSPREVIHGRALGRVAPSWLGLSLSEALCFGLATLPMNAVTPALTSGLAYIL 782  
 QY 783 DFLQMSAFVALILSDSRKQASRLDVCCVYKQELPPPGQBGILLGFFOKAVAPFLH 842  
 DB 783 DFLQMTAFVALILSDSRKQASRDVCCFSRRLPPPKQEGILLGFFPRKIVYFLH 842  
 QY 843 WITRGVILLPLAFGVSLYKCHISVGLDQELAPKQSYLDYFLFNRIFEVGAHYE 902  
 DB 843 RIRRVVILLFLVLFGANLYMCNISVGLDDDLAPKQSYLDYFLFNRIYLEVGPYIF 902  
 QY 903 VTTGYNSSREGMNAICSSAGCNNEFTOKIOVATEPBGOSYALIPASSWVDPIWLT 962  
 DB 903 DTTSYNPSSTEGMNAICSSAGCSFSLTQKIQVASEFPNOSYVALIASWVDPIWLT 962  
 QY 963 P-SSCCRLYISGPNKDKCPSVANSINCLAKCMISITMGSVBPVQFHKYLPWFINDRPN 1021  
 DB 963 PSSSCRLYITRPHNDEFCEPSTDISFNCLKCMNRITLCPVPTTEQFHKYLPWFINDRPN 1022  
 QY 1022 IKCPKGLAAVSTSNLTSDQVTLASRFMAVHKPLKNSQDYTBALBAARELANITADLR 1081  
 DB 1023 IRCPKGLAAVSTSNLTSDQVTLASRFMAVHKPLKNSQDYTBALBAARELANITADLR 1082  
 QY 1082 KVPGTDPAPFVFPYITNVFBOYVTLIPBGLFPLMLSLCLVFPFAVSCLLGLDLSGLIN 1141  
 DB 1083 KVPGTDPAPFVFPYITNVFBOYVTLIPBGLFPLMLSLCLVFPFAVSCLLGLDLSGLIN 1142  
 QY 1142 LLSIVMILVDTVGFMAIDISNAVSLINLVSAVMSVEFVSHITRSPALSTKPTMLERA 1201  
 DB 1143 LLSIMILVDTVGLMAVGVISVNAVSLINLVTAVMSVEFVSHITRSPAVSTKPTMLERA 1202  
 QY 1202 KEATISGSAVPAVAMTNLPGILVGLAKAQLOIIFPRNLNLTLLGLHGLVFLPV 1261  
 DB 1203 KQATIFMSAFAVAVAMTNLPGIILGFAQAOLOIIFPRNLNLTLLGLHGLVFLPV 1262  
 QY 1262 LSYGPDNPALALEQKRAEEVAAVMAASCPNHSRSTADNTIYVNSPEGS-1KGAGA 1320  
 DB 1263 LSYGPDNPQALVLEKQATEA-AWVSEPPSCQYFPADANTSDVYVNGFNPFIPEINA 1321  
 QY 1321 ISNPLPNNGROP 1332  
 DB 1322 ASSSLPKSDQKF 1333

RESULT 15  
 ADY60748 ID ADY60748 standard; protein; 1333 AA.  
 XX  
 AC ADY60748;  
 XX  
 DT 19-MAY-2005 (first entry)  
 XX  
 DE Mouse Niemann-Pick C1-like protein 1 (NPC1L1) protein.  
 XX  
 KW Niemann-Pick C1-like protein 1; Antihypemic; Antiarteriosclerotic;  
 KW Cardiant; Vasotrophic; Cerebroprotective; cholesterol; hyperlipidemia;  
 KW metabolic disorder; atherosclerosis; cardiovascular disease;  
 KW coronary artery disease; cerebrovascular ischemia; neurological disease;  
 KW arteriosclerosis; transgenic animal.  
 XX  
 OS Mus sp.

XX WO2005015988-A1.  
PN  
XX  
XX 24-FEB-2005.  
PD  
XX  
XX 16-DEC-2003; 2003WO-US040113.  
PF  
XX  
XX 17-JUL-2003; 2003US-00621758.  
PR  
XX 22-AUG-2003; 2003US-00646301.  
PR 16-SEP-2003; 2003US-00663208.  
XX  
XX (SCHER) SCHERING CORP.  
PA  
XX  
XX Altman SW, Murgolo NJ, Wang LQ, Graziano MP;  
PI  
XX  
XX MPI; 2005-284403/29.  
DR  
XX N-PSDB; ADY60747.  
DR  
XX  
XX Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of  
PT NPC1L1 that inhibits NPC1L1 mediated sterol or 5 alpha-sterol uptake, for  
PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.  
XX  
XX Claim 30; SEQ ID NO 12; 146pp; English.  
XX  
XX The invention relates to an isolated Niemann-pick C1-like protein 1  
CC (NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an  
CC amino acid sequence appearing as ADY60738, ADY60740 or ADY60748 (Rat,  
CC human and Mouse NPC1L1). Also included are an isolated polynucleotide  
CC encoding NPC1L1 (appearing as ADY60737, ADY60739 or ADY60747), a  
CC recombinant vector comprising the polynucleotide, a host cell comprising  
CC the vector, an isolated antibody which specifically binds to NPC1L1 (or  
CC to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),  
CC preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous  
CC mutation of endogenous chromosomal NPC1L1 (where the mouse does not  
CC produce any functional NPC1L1 protein), an offspring or progeny of the  
CC mouse having inherited a mutated NPC1L1 allele of the mouse, screening a  
CC sample for an intestinal sterol or 5alpha-sterol absorption antagonist,  
CC inhibiting NPC1L1 mediated sterol or 5alpha-sterol uptake in a subject, a  
CC kit (comprising ezetimibe or substituted azetidinone in a pharmaceutical  
CC dosage form, and information indicating that NPC1L1 is a target of  
CC ezetimibe or substituted azetidinone), decreasing the level of intestinal  
CC sterol or 5alpha-sterol absorption in a subject (involving reducing the  
CC level of expression of NPC1L1 in the subject), identifying an antagonist  
CC of NPC1L1 and an isolated mammalian cell which lacks a gene encoding a  
CC functional NPC1L1 protein. NPC1L1 is useful for identifying an antagonist  
CC of NPC1L1 which is useful for inhibiting or decreasing the level of  
CC NPC1L1 mediated sterol or 5alpha-sterol uptake in a subject such as  
CC mouse, rat or human. The level of expression of NPC1L1 in the subject is  
CC reduced by mutating NPC1L1 in the subject. The NPC1L1 knockout mouse is  
CC useful for screening a sample for intestinal sterol or 5alpha-sterol  
CC absorption antagonist. The NPC1L1 antagonists identified are useful for  
CC treating medical conditions caused or mediated by NPC1L1, e.g.,  
CC hyperlipidemia, atherosclerosis, coronary heart diseases, stroke or  
CC arteriosclerosis. The present sequence represents a mammalian NPC1L1  
CC protein.  
XX  
XX  
SQ Sequence 1333 AA;  
Query Match 78.3%; Score 5407; DB 9; Length 1333;  
Best Local Similarity 77.2%; Pred. No. 0;  
Matches 1028; Conservative 126; Mismatches 174; Indels 4; Gaps 4;

QY 183 LAVGTCGVYGSALCNAQRWLNFGDGTGNGLAELDITPHLLFEGQAVSGIOPLENGVAR 242  
183 LAVGTCGVYGSALCNAQRWLNFGDGTGNGLAELDITPHLLFEGQAVSGIOPLENGVAR 242  
DB 183 LAVGTCGVYGSALCNAQRWLNFGDGTGNGLAELDITPHLLFEGQAVSGIOPLENGVAR 242  
QY 243 CNEGQDDVATCGCQCCAAACCPAIARPOLDSFYLGOMGSGVLIIILCSAVAVTIIIL 302  
243 CNEGQDDVATCGCQCCAAACCPAIARPOLDSFYLGOMGSGVLIIILCSAVAVTIIIL 302  
DB 243 CNEGQDDVATCGCQCCAAACCPAIARPOLDSFYLGOMGSGVLIIILCSAVAVTIIIL 302  
QY 303 VGFPRVAPARDKSGMVPDKGTSLSDDLSTFTHLLLOFQCGMGWVAVWPLTIIIVLSVIP 362  
303 VGFPRVAPARDKSGMVPDKGTSLSDDLSTFTHLLLOFQCGMGWVAVWPLTIIIVLSVIP 362  
DB 303 VGFPRVAPARDKSGMVPDKGTSLSDDLSTFTHLLLOFQCGMGWVAVWPLTIIIVLSVIP 362  
QY 363 VVALAGLVFTTELTTDPVEIWSAPNSQASEKAFFHDQHPFPRITNOVILTAIPRSSRY 422  
363 VVALAGLVFTTELTTDPVEIWSAPNSQASEKAFFHDQHPFPRITNOVILTAIPRSSRY 422  
DB 363 VVALAGLVFTTELTTDPVEIWSAPNSQASEKAFFHDQHPFPRITNOVILTAIPRSSRY 422  
QY 423 DSIILGPKNFGSGLIDLDLLELEQLERLHLQWSPBEAQRNLSLDICTAPLNPDTSL 482  
423 DSIILGPKNFGSGLIDLDLLELEQLERLHLQWSPBEAQRNLSLDICTAPLNPDTSL 482  
DB 423 DSIILGPKNFGSGLIDLDLLELEQLERLHLQWSPBEAQRNLSLDICTAPLNPDTSL 482  
QY 483 YDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGTAALASC 542  
483 YDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGTAALASC 542  
DB 483 YDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGTAALASC 542  
QY 543 MADYGA PVFEPFLAIGGYKGYSEABALIMTFSLANNYPADGPRLAQAALWEBAFLBEMRA 602  
543 MADYGA PVFEPFLAIGGYKGYSEABALIMTFSLANNYPADGPRLAQAALWEBAFLBEMRA 602  
DB 543 MADYGA PVFEPFLAIGGYKGYSEABALIMTFSLANNYPADGPRLAQAALWEBAFLBEMRA 602  
QY 603 FQRMAGMFOVTTAERLSLEDEINRTTAEDLPFATSYIYIPLYISALGSSYSWSVMV 662  
603 FQRMAGMFOVTTAERLSLEDEINRTTAEDLPFATSYIYIPLYISALGSSYSWSVMV 662  
DB 603 FQRMAGMFOVTTAERLSLEDEINRTTAEDLPFATSYIYIPLYISALGSSYSWSVMV 662  
QY 663 DSKRTTGLGVAIVLVANMAAMGFYIGRSSVLVLOVVPVLVSAGADNIFLVEYQ 722  
663 DSKRTTGLGVAIVLVANMAAMGFYIGRSSVLVLOVVPVLVSAGADNIFLVEYQ 722  
DB 663 DSKRTTGLGVAIVLVANMAAMGFYIGRSSVLVLOVVPVLVSAGADNIFLVEYQ 722  
QY 723 RLPRRGEPEPEVHIGRALGVAPSMILCSLEAICEFLGATMPAVRTALTSGLAVIL 782  
723 RLPRRGEPEPEVHIGRALGVAPSMILCSLEAICEFLGATMPAVRTALTSGLAVIL 782  
DB 723 RLPRRGEPEPEVHIGRALGVAPSMILCSLEAICEFLGATMPAVRTALTSGLAVIL 782  
QY 783 DFLQWAFVALLSLDSKROEASRLDYCCVQDELPPEQSGSLIGFQKAYAPRLH 842  
783 DFLQWAFVALLSLDSKROEASRLDYCCVQDELPPEQSGSLIGFQKAYAPRLH 842  
DB 783 DFLQWAFVALLSLDSKROEASRLDYCCVQDELPPEQSGSLIGFQKAYAPRLH 842  
QY 843 WTRGVVLLFLALFGVSYLSMCHISVGLDDELALPDVSLDYFLFANNYPFEGAVYF 902  
843 WTRGVVLLFLALFGVSYLSMCHISVGLDDELALPDVSLDYFLFANNYPFEGAVYF 902  
DB 843 WTRGVVLLFLALFGVSYLSMCHISVGLDDELALPDVSLDYFLFANNYPFEGAVYF 902  
QY 903 VTTGYNFSEAGMNAICSSAGCNFSFTQKIQATEFPQSYLAI PASSWVD FIDMLT 962  
903 VTTGYNFSEAGMNAICSSAGCNFSFTQKIQATEFPQSYLAI PASSWVD FIDMLT 962  
DB 903 VTTGYNFSEAGMNAICSSAGCNFSFTQKIQATEFPQSYLAI PASSWVD FIDMLT 962  
QY 963 P-SSCCRLYISGNKDKFCBSTVNSLNCNKNCSITWGSVRSVBOFHXYLPMFLNDPN 1021  
963 P-SSCCRLYISGNKDKFCBSTVNSLNCNKNCSITWGSVRSVBOFHXYLPMFLNDPN 1021  
DB 963 P-SSCCRLYISGNKDKFCBSTVNSLNCNKNCSITWGSVRSVBOFHXYLPMFLNDPN 1021  
QY 1022 IKCPKGLAAYSTSVNLTSDGVLASRFMAVYHPELKNQSDPYTALPARELANITADLR 1081  
1022 IKCPKGLAAYSTSVNLTSDGVLASRFMAVYHPELKNQSDPYTALPARELANITADLR 1081  
DB 1022 IKCPKGLAAYSTSVNLTSDGVLASRFMAVYHPELKNQSDPYTALPARELANITADLR 1081  
QY 1082 KVPETDPAFVFPYTTTNNVYBOYLITLPGFLFMLSCLVPTPAVSCILGLDLRSGLN 1141  
1082 KVPETDPAFVFPYTTTNNVYBOYLITLPGFLFMLSCLVPTPAVSCILGLDLRSGLN 1141  
DB 1082 KVPETDPAFVFPYTTTNNVYBOYLITLPGFLFMLSCLVPTPAVSCILGLDLRSGLN 1141  
QY 1142 LLSIWMILVDTVGFMAIMDISYNAVSLINIVSAVNGSVFVSHITRSFAISTYKPTWLER 1201  
1142 LLSIWMILVDTVGFMAIMDISYNAVSLINIVSAVNGSVFVSHITRSFAISTYKPTWLER 1201  
DB 1142 LLSIWMILVDTVGFMAIMDISYNAVSLINIVSAVNGSVFVSHITRSFAISTYKPTWLER 1201  
QY 1202 KEATISGSAVFPAGVAMTNIPGLIVUGLAKAOILQIFFFFLNLTITLGLHGVFLPVI 1261  
1202 KEATISGSAVFPAGVAMTNIPGLIVUGLAKAOILQIFFFFLNLTITLGLHGVFLPVI 1261  
DB 1202 KEATISGSAVFPAGVAMTNIPGLIVUGLAKAOILQIFFFFLNLTITLGLHGVFLPVI 1261  
QY 1203 KDATIFMGSAVFAGVAMTNIPGILIGFAOAOILQIFFFFLNLTITLGLHGVFLPVI 1262  
1203 KDATIFMGSAVFAGVAMTNIPGILIGFAOAOILQIFFFFLNLTITLGLHGVFLPVI 1262  
DB 1203 KDATIFMGSAVFAGVAMTNIPGILIGFAOAOILQIFFFFLNLTITLGLHGVFLPVI 1262

Qy	1262	LSYVGPDVNPALALEQKRAEEVAAMVNASCPNHPSEVSTADNITYNHSFEGS-1KGAG	1320
Db	1263	LSYLGPDVNOALVLEKLA TEA-AMVSESPSCPOYPFPADANTSDYVNYGFFNPEFIPEINA	1321
Qy	1321	ISNPLPNNGROF	1332
Db	1322	ASSSLPKSDQKF	1333

Search completed: March 22, 2006, 22:50:48  
Job time : 209 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 22:55:32 ; Search time 52 Seconds  
(without alignments)  
2117.769 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909  
Sequence: 1 MAEAGLRGWLIMALLRLAQ.....GSIKGAGAISFLPNNGRQF 1332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : Issued Patents AA:*
1: /cgn2.6/prodata/1/1aa/5 COMB.pep:*
2: /cgn2.6/prodata/1/1aa/5 COMB.pep:*
3: /cgn2.6/prodata/1/1aa/R COMB.pep:*
4: /cgn2.6/prodata/1/1aa/PTURS COMB.pep:*
5: /cgn2.6/prodata/1/1aa/RE COMB.pep:*
6: /cgn2.6/prodata/1/1aa/backfill1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2403	34.8	1318	2	US-09-949-016-10152	Sequence 10152, App1
2	2402.5	34.8	1278	2	US-09-462-136-2	Sequence 2, App1
3	2385	34.5	1319	2	US-09-462-136-4	Sequence 4, App1
4	1329	19.2	1170	2	US-09-462-136-6	Sequence 6, App1
5	1078	15.6	1296	2	US-09-462-136-9	Sequence 9, App1
6	651.5	9.4	1447	1	US-08-540-056-19	Sequence 19, App1
7	651.5	9.4	1447	1	US-08-656-055-19	Sequence 19, App1
8	651.5	9.4	1447	2	US-08-954-668-19	Sequence 19, App1
9	651.5	9.4	1447	2	US-09-268-140-5	Sequence 5, App1
10	651.5	9.4	1447	2	US-08-918-658-19	Sequence 19, App1
11	651.5	9.4	1447	2	US-09-724-631-19	Sequence 19, App1
12	651.5	9.4	1447	2	US-08-954-701A-19	Sequence 19, App1
13	651.5	9.4	1447	2	US-09-807-007-6	Sequence 6, App1
14	651.5	9.4	1447	2	US-09-754-032-19	Sequence 19, App1
15	651.5	9.4	1447	2	US-08-916-140-19	Sequence 19, App1
16	651.5	9.4	1447	4	PCT-US95-13333-19	Sequence 19, App1
17	651.5	9.4	1434	1	US-08-540-406-10	Sequence 10, App1
18	646.5	9.4	1434	2	US-08-656-055-10	Sequence 10, App1
19	646.5	9.4	1434	2	US-08-954-668-10	Sequence 10, App1
20	646.5	9.4	1434	2	US-08-918-658-10	Sequence 10, App1
21	646.5	9.4	1434	2	US-09-724-631-10	Sequence 10, App1
22	646.5	9.4	1434	2	US-08-954-701A-10	Sequence 10, App1
23	646.5	9.4	1434	2	US-09-754-032-10	Sequence 10, App1
24	646.5	9.4	1434	2	US-08-916-140-10	Sequence 10, App1
25	646.5	9.4	1434	4	PCT-US95-13233-10	Sequence 10, App1
26	639.5	9.3	642	2	US-09-246-796A-16161	Sequence 16161, App1
27	628	9.1	1296	2	US-08-857-636-60	Sequence 60, App1

28	584	8.5	1203	2	US-09-2297-505-2	Sequence 2, Appl 1
29	584	8.5	1203	2	US-09-060-939A-2	Sequence 2, Appl 1
30	583.5	8.4	1182	2	US-09-231-505-7	Sequence 7, Appl 1
31	583.5	8.4	1182	2	US-09-060-939A-7	Sequence 7, Appl 1
32	581.5	8.4	531	2	US-09-248-786A-16160	Sequence 16160, A
33	580	8.4	1203	2	US-09-207-857-2	Sequence 2, Appl 1
34	580	8.4	1203	2	US-09-909-280A-2	Sequence 2, Appl 1
35	579	8.4	1146	2	US-09-807-007-1	Sequence 1, Appl 1
36	526.5	7.6	1311	1	US-08-540-406-4	Sequence 4, Appl 1
37	526.5	7.6	1311	2	US-08-656-055-4	Sequence 4, Appl 1
38	526.5	7.6	1311	2	US-08-954-668-4	Sequence 4, Appl 1
39	526.5	7.6	1311	2	US-08-918-658-4	Sequence 4, Appl 1
40	526.5	7.6	1311	2	US-09-724-631-4	Sequence 4, Appl 1
41	526.5	7.6	1311	2	US-08-954-701A-4	Sequence 4, Appl 1
42	526.5	7.6	1311	2	US-08-925-032-4	Sequence 4, Appl 1
43	526.5	7.6	1311	2	US-08-918-140-4	Sequence 4, Appl 1
44	526.5	7.6	1311	4	PCT-US95-13233-4	Sequence 4, Appl 1
45	474	6.9	1286	2	US-09-266-140-3	Sequence 3, Appl 1

## ALIGNMENTS

```

RESULT 1
US-09-949-016-10152
/ Sequence 10152, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10152
/ LENGTH: 1318
/ TYPE: PRT
/ ORGANISM: Human
US-09-949-016-10152

Query Match      34.8%; Score 2403; DB 2; Length 1318;
Best Local Similarity 39.7%; Pred. No. 3,9e-218;
Matches 525; Conservative 238; Mismatches 451; Indels 110; Gaps 26;

QY      2 AEAGLRGULMA-----LLRLAQSSEPYTTIHOPGYCAFYDECG-----KNPELGS 47
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      32 ASALRGHSMARGLALGLLLLLLCPAOVFSQ-----SCWVGEGCIAYGDKRKYCEYSG 86
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      48 SLMTLSNVSCLSNTPARKITGDHLLTLKICRGLTYTPPTQACCSAKOLVSLSEASLIRK 107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      87 -----PPKPLPKDGYDLVDELCEGFFFG-NVSLCCDVRQLQTLKNDLQILPL 131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      108 ALITFCPCSDNFVNLHGNTCSPNQSLFIYNT-----VAOLGAGQLPAVVAIEAFYQHS 163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      132 QFLSBCPCFYNLLNLFCELTCSPRQSGFLNVTATEDYVDPTNQTIVKELOXYVVGQS 191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      164 PAEGSYDSCSRVRVAATTLAVGTMGCGYVGSALCNAAQKMLNQGDTGNGLAPLDT----- 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      192 PANAMYNACRDEAVSESSNDKALGLLGCXDAAD-CNATWIEEMFNKNDQAPPTITPVFS 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      220 ---FILEPQAVGSGEIOPLNEGVARCNESQGDVATCSQDCAASCFAIARPO----- 270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      251 DPVPH-----GMEPANNATKGCBSDEVETAPSCQDQCSIVCGPKPQPPPPAPM 300
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      271 ---ALDSTFVYGQMGESLVLIIILCSVERAVTITL-----VGRVAPARQDS 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 301 TILGLDAMVIMITYMAFLVFFGAFVWVCYRKRKYFVSEYTPIDISNTAFSV-NASDGG 359  
Qy 315 KMWDEKGTSLSDKLSFTHTLLGQFFQMGTMVASMPLTIIVLVSVVLAALGVLTE 374  
Db 360 E-----ASCDDPVSAAFEGCLRLFTRMGSFCVRNPPCVIFPSLVFTTACSSGLVFR 412  
Qy 375 LTTDPVELMSAPNSQARSKAHFDHGFPPFTNOVILTAPNRSRYRSDLLGPK-NFS 433  
Db 413 VTTNPDVMSAPSSQARLEKEYFDQHFGBFFRTEQLIRAPLTDKHIYQRPESGADVPG 472  
Qy 434 GILDDLLELELQERLRHLQVMSPEAQRNISLQDICAPLNPDNTSLYDCCINSLOY 493  
Db 473 PPLDIOIHLQVLDQIAIEN--ITASYDNETVTLQDICAPISPVNT--NCTIISLVNY 527  
Qy 494 FQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKQDGTALALSCMADYAVPFP 553  
Db 528 FQNSHSVLDHKKGDDEF--VYADHYTHFLYCVARASLNDTSLHDPCLGTFGGVFPW 584  
Qy 554 LAIGYKGDYSBAEALIMTFSLNNYPADPRLAOKLWEAFLBEMRAFORRMAGMFGV 613  
Db 585 LVLAGYDDQNNYNAATLVITPPVNNYNDTEKQRAQWKEKFINVKYKN--PNLTI 641  
Qy 614 TTTARSLDEINRTTAEPLPFATSYIYIFLYISALGSYSKRWMDSKATIGLGV 673  
Db 642 SFTARSLDEINRSDSDVFTVVISYALMFLYISLALGHKSCRLVDSKVSIGIAI 701  
Qy 674 AVVLGAVMAAMGFSLGIRSLVLQVPPFLVLSVADNIFFLVLEYORLRRPEPE 733  
Db 702 LIVLSSVACSLGVFSYIGPLTIVIEVPLVLANGVNIPLVQAYORDERLQGETID 761  
Qy 734 VHIQALGRVAPSMLLCSLSEAI CFFLGLATMPAVRTALTSGLAVILDFLLQNSAFYA 793  
Db 762 QQLGVILGAVASMFSSFSFETVAFPLGALSVMPAVHTSLFAGLAVFIDFLQITCFVS 821  
Qy 794 LILSLSKROEASRLDYCCVQPELPPGO--GBGILLGFQKAYAPFLIMLITRQVLL 852  
Db 822 LILGLIKROEKRLDIPCCVRAGDGTSVQASBSCIFRFKNSYSPLLLKDMRPVIAI 881  
Qy 853 FLATFVSLYSMCHISVGLDOELALPKDSYLLDYFLNRYFEVGAFFVTTLLGYNFS 912  
Db 882 FVGVLSSFIAVLNKKDIDGLDOSLMDSDYMDYFYSISQYLHAGPPVFLVEBGHDYS 941  
Qy 913 EAGMNAICSSAGCNNSFTQKI QYATEPEQSYLAI PASSWDDFLWLT-SSCCRLXI 971  
Db 942 SKGQNVVCGMGCCNNDLVOQIFNAQQLDNYTRIGFAPSMDIDYFDVKPOSSCCRY-- 999  
Qy 972 SGPNKDKFCPSYVNSLCLNCKMSTI-MGSVNPVSVEQFHKYLPWFLANDPNIKCPKGLA 1030  
Db 1000 -DNITDQFCNAAVVPACVR-CRPLPBEKQRPQGGDFMRFLPMFLSDNPNKCGKGHA 1057  
Qy 1031 AYSTSVN--LTSQGVLAGRFMAVHKPLKNSODYEALRAARELANITADIRKVGTP 1088  
Db 1058 AVSSAVNILLGCTRGVATYFMTHTVLOTSDAFIDALKKARLIASNT-ETMGINGS-- 1114  
Qy 1089 AEEVEPYITTVNYEBOYLTIPEGLFMLSCLVTPFVAVSCLLGLDRLSGNLNLISYVI 1148  
Db 1115 AAVPEPYAVFYEBQYLTIIDTIPNLGVSIGAFIVTMVLLGCLMGAIVIMCATIANV 1174  
Qy 1149 LVDYGFMAIINDISYXAVSLINLVSAVMSVEFVSHITSFPAISTPTMLERAKERTIM 1208  
Db 1175 LVNMGMVWLWGISLNAVSLVNLVMSCGISVEFCSHITRAFTVSMKGSVEBAEELAHM 1234  
Qy 1209 GSAVAGVAMTNLPGLIVLGLAKAOLQIFFPRLNLLITLGLLHGLVFLPYLSSVGP 1268  
Db 1235 GSSVSGITLTKRGGIVLAPAKSOLFQIFPRMYLAMLGATHGLIFLPVLSITGFS 1294  
Qy 1269 VNPA 1272  
Db 1295 VNKA 1298

US-09-462-136-2  
/ Sequence 2, Application US/09462136  
/ Patent No. 6426198  
/ GENERAL INFORMATION:  
/ APPLICANT: Carleia, et al.  
/ TITLE OF INVENTION: Genes for Niemann-Pick Type C Disease  
/ FILE REFERENCE: 4239-53894  
/ CURRENT APPLICATION NUMBER: US/09/462,136  
/ CURRENT FILING DATE: 2000-06-01  
/ PRIOR APPLICATION NUMBER: PCT/US98/13862  
/ PRIOR FILING DATE: 1998-07-02  
/ PRIOR APPLICATION NUMBER: US 60/051,682  
/ PRIOR FILING DATE: 1997-07-03  
/ NUMBER OF SEQ ID NOS: 13  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 2  
/ LENGTH: 1278  
/ TYPE: PRN  
/ ORGANISM: Homo sapiens  
US-09-462-136-2  
Query Match 34.8%; Score 2402.5; DB 2; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 4.2e-218;  
Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;  
Qy 7 RGMILMALRLAQSEPYTTIHQPGYCAFYDECG-----KNPELGSGLMTLSNVCLIS 59  
Db 4 RGLALGLLTLTLCPAQVFSQ-----SCWYGECCGAYGDKRYNCYESG----- 46  
Qy 60 NTPAKITGDHILLOKICRPLHYGNPNQACCSAKOLVLSIASITLALITRCPACSDN 119  
Db 47 --PPPLPDGYDVLQVLCPEFFG-NVSLCCDVQVQTLQKNDLQPLQFSRCSFYRN 103  
Qy 120 FVNLCHNTGSPNQLFTINVR---VAQAGAGQLPAAVAAEAFYQHSFAEQSYDSCSRV 175  
Db 104 LNLCELTCSRQSQFLNTATBEDYDPTVNTQRTVKNELQYVYGQSPANMYAACDV 163  
Qy 176 RVPAAATLAVGTGCVGSALCNAQRWLNFGQDGTNGIAPLDT-----FHLEPQQA 228  
Db 164 EAPSSNDALGLLCSKADA-CNATNMEIYMNKNQGAFFITVPFSDFPVH----- 215  
Qy 229 VQSGIQLNBEVARGNESQGDVATCSGODCAASGPALARP-----ALDSTFYLG 279  
Db 216 ---GMEPMNNAATKGCDESVDVETAPCSCODCSIVCGPKPQPPPPAPMTIIGLDMVYIM 272  
Qy 280 QWPGSLVLIILICSVFAVVTILL-----VGFVAPARDAKAMVDPKKGTSLS 326  
Db 273 WITTYAFLVFFGAFVWVCYRKRKYFVSEYTPIDISNTAFSV-NASDGE-----ASC 324  
Qy 327 DKLSTFTLLGQFFQMGTMVASMPLTIIVLVSVVLAALGVLTELTDPVELMSAP 386  
Db 325 DVSAAAFEGCLRLFTRMGSFCVRNPPCVIFPSLVFTTACSSGLVFRVTTNPDVMSAP 384  
Qy 387 NSQASSEKAFHDHGFPPFTNOVILTAPNRSRYRSDLLGPK-NFGIILDDLLELL 445  
Db 385 SQQARLEKEYFDQHFGBFFRTEQLIRAPLTDKHIYQRPESGADVPEPPLDIOIHLQV 444  
Qy 446 ELQERLRHLQVMSPEAQRNISLQDICAPLNPDNTSLYDCCINSLOYFQNNRTLLTLTA 505  
Db 445 DQIAIEN--ITASYDNETVTLQDICAPISPVNT--NCTIISLVANFQNSHSLVDHKK 499  
Qy 506 NOTLMGQTSQVDMKDHFLYCANAPLTFKQDGTALALSCMADYAVPFPPLAIGYKGDYS 565  
Db 500 GDDPF--VYADHYTHFLYCVARASLNDTSLHDPCLGTFGGVFPWFLVGLGYDDQVYN 556  
Qy 566 EAEALIMTFSLNNYPAGRPRLAOKLWEAFLBEMRAFORRMAGMFGVOTFTTERLEDEI 625  
Db 557 NATALVITPPVNNYNDTEKQRAQWKEKFINVKYKN--PNLTIISFTARSLIEBL 613  
Qy 626 NRTTAEPLPFATSYIYIFLYISALGSYSKRWMDSKATIGLGVAVVLGAVMAAMG 685  
Db 614 NRESDDVFTVVISYALMFLYISLALGHKSCRLVDSKVSIGIAGLILVLSVACSLG 673

```

Qy      686 FFSVLGIRSSVLIIQVAVFVLVSAGADNIFFLFLVLYQRLPRRPGEPREVHIGRLAKRVP 745
Db      674 VFSYIGLPLTLTIVLEIPIFLVAVGVNDI FLTVAYQGRDETLQGETLDDQGLRGVAGEVAP 733
Qy      746 SMLCSLSEALCFPLGALTTPMPAVRTFLTJSGLAVIDFLQOMAFVALLSLDKRDEAS 805
Db      734 SMLFSFSEYVAFVFGALSVMPAVHTSFLAGLVAFIDFLQITCFPSLGLDKRDEKN 793
Qy      806 RLVDCCCKRPELPPRPG-GEGLLGGFQKAYAPFLHWMIRGVVLLFLFLFVSLYSM 864
Db      794 RLDFPCCVRGABDGTSVQASBSCFLRFPKNSYSPLLKDMNRPLVIALFVGLSFSFAVL 853
Qy      865 CHISVGLDQELALPKDSYLLDYLFLFNKYFEVGA PVYFVTLTGYNFSSEAGMALICSSAG 924
Db      854 NKVIGLQDLSMPDSDSYMVDYFKSISQYLHAGPPVYFVLEEGHDYTSKQNNVCGMG 913
Qy      925 CNNSFTOKIQYATFEPQSYIATVPASSWVDDEFIDMLTP--SSCCRLYISGPNKDFCEPST 983
Db      914 CNNDSLVQOITNMAQLDNYTRIIGRAPSSWIDDYFDMVAPQSSCCRV--DNITQFCMAS 970
Qy      984 VNSLNCLEKNCMSIT-MGSVRPSVEQFHKYLEPFLNDRPNICCPKXGLAAYSTVN--LTS 1040
Db      971 VVDPRACVA-CRPLTPREGKQRPQGGDFMFLFEMFLSDNPNPCGKGHAHAYSAVANIILGH 1029
Qy      1041 DGOVLASRFMAVYHKPLKNSODYTEALRAARELANITADLRKPGTDPAFVFPYITNV 1100
Db      1030 GTRVAGATFMYHYTLQTSADPFDLKKARILASVLT-ETMINGS--AYRVFPYSVYV 1086
Qy      1101 FVEQYLTLLPBGFLPMLSLCLVPFAVSCLLGLDLSGLNLNLSIMVLVDTVGPMALMD 1160
Db      1087 FVEQYLTLLIIDPTIFNLGVSGLAIFLVTVWLLGCELSMAVICATIAVLVNMFGVWMLMG 1146
Qy      1161 ISYNAVSLINLVSAVGMSEVFVSHITRSPALISTKPTWLERAKERTISMGSVAPGVAMTN 1220
Db      1147 ISLNAVSLINLVMSGCSIVSEFCSHITRAFTVSMKGSRYERABEALAHMGSSVFSGITLT 1206
Qy      1221 LPLGIVGLAKAQIQTFFPRLNLTITLGLLHGVLPLVLSYVGPVNNNA 1272
Db      1207 FGGIVVLAFAKSQIPQIIFYFRMYLAVMLLGAATHGLIFLPLLSYIGPENVKA 1258

RESULT 3
US-09-462-136-4
; Sequence 4, Application US/09462136
; Patent No. 6426198
; GENERAL INFORMATION:
; APPLICANT: Carstea, et al.
; TITLE OF INVENTION: Genes for Niemann-Pick Type C Disease
; FILE REFERENCE: 4239-53894
; CURRENT APPLICATION NUMBER: US/09/462,136
; CURRENT FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: PCT/US98/13862
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/051,662
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1319
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-462-136-4

```

	Query Match	34.5%	Score 2385,	DB 2,	Length 1319;
	Best Local Similarity	40.0%;	Pred.No.	2e-216;	
	Matches	524;	Conservative	231; Mismatches	440; Indels 116; Gaps 28;
Oy	14 LLLRLAOSPPYTIHQPGCAFYDCGKGNPELGSIMLTLSNVSCLSNTPARKITGDHIL	73	:	:	:
Dd	53 LLHLCPAQVFSQ-----SWWYGEGC---LATGP-----KRNCCKXSGEPKPDLPRDGVLD	100	:	:	:
Oy	74 LOKICRPLYTGENTQAACSAKOVLSEASLSITKALLTRPCAGSDNFVNLIHCMTCSNQ	133	:	:	:

Db	101	VQELCPGLFF-DNVSLOCDDIQLOTLKSNQLPLQFTRSCBSCRYNMTJFCBLTCSPHQ	159
Qy	134	SLFINTVRVAOLGAGOLPA-----VVAEAFYQHSFAQSDSCSRVRVPAATLAVGTMC	189
Db	160	SQFLNATATEDYFDPKTPENKTNVKELEYVGOSFAANMYNA-----CRADEAPSSNEKALGLIC	219
Qy	190	GTVSSALCNACRMLNFGODTGNGLAPLDI-----THLLBEGQAVGGSIQPLNBGVARN	244
Db	220	GRDARA-CNATNMWIEFMFNKMDNGAPFTIIPVPSDSL-----GMEBMRNATKCN	270
Qy	245	ESQGDVATGSCQDCOASCALARPQMLDSTFYLGQMPGSLVLIILCSVAVVTTILLG	304
Db	271	ESVDEVTPGSCDCDCSIVCG--PRPOLLQIGGGGGM-GUDAMVIMVTVVAFLEVPFG	327
Qy	305	FRVA-----PARSKXMDVPKKGTSLSPDKLSPSTHTLLGQF	340
Db	328	ALLAVMCHRRRYFYSEYTPIDSNIAFSVNSDKGE-----ASCCPLGAAPDCLRRM	380
Qy	341	FOGKGTWASWPLTILVSVIPVVALAAGVFTLTTPVELMSPNSOARSEKAFDHOH	400
Db	381	FTKGAACVARNPTCIIFFSLAFIVCGSGLVQVYTNIPVELMSPNSQARLEKEYPDXH	440
Qy	401	FGPFRTNOVILTAPNRSSYRYDSSLGPK-NFSGIILDLLELLELBERLRLQVWSP	459
Db	441	FGPFRTREQILQAPNPSVHIYEYEPAGADVFGPPLNKETILHQVNLQ-----IAIESI	495
Qy	460	EAGQN---ISLQDTCVAPLPNDNTSLYDCCNSLQYFQNNRTLLLTANQTLMGQTSQV	516
Db	496	TASYNNEVTVLQDICIAPLSPYNK--NCTIMSYLANFONSHAUL-----DSQV	541
Qy	517	-----DWKDHFLXCANPLTKDQDTALALSGMADYGAPVPEFLAIGYKGDQVSAE	568
Db	542	GDDFYIADYHTEFLCVAPASINDTSLHGPCLTGFGGVPFWLVAGDDQNYNAT	601
Qy	569	ALINTFSLNNYPAGDPRLOAKLMEBAFLBEMARFORRMAGMFOVTEAERLDEINRT	628
Db	602	ALVTTFPPNNYNDTEKLRQAMAMEKEFISVAKYKA---PNLTISTAEISIDELNRE	658
Qy	629	TAEDLPFATSYIIFLYISLALGSYSWSRWMDSKATLGLGAVAVVLAGVMAAMEFFS	688
Db	659	SNSDVFYTIISYVMFLYISLALGHQSCSLVDYSKISLGTAGILLVSSVACSLGIFS	718
Qy	689	YLGRSSLVILQVVPFLVLSVGNDFIIVLEYORLPRRGEPREVHIGRALGVASBML	748
Db	719	YMGPLTFLIEVPIPLVLAGVDNIFLVQTYQRDEBLQETDQQLGILGEVAPATMF	778
Qy	749	LCSISEAICFPLGALTMPAVRTALTSGLAVIIDFLQNSAFALLSLDSKROBASLD	808
Db	779	LSSFSEISAFPFGLASSMPAVHTSLPAGMAVILIDFLQITCFVSLTGLDIKROEKMLD	838
Qy	809	VCCCVKQOELPPPOG-----EGLLGFQKAYAPFLHMTIRGVULLFLALFQVLSYM	864
Db	839	ILCCVRGAD--DOQSHASESYLFRFFKYNFAPLLKMDLPIVVAAPGVLSFSVAUV	895
Qy	865	CHISVGDOELAKRDSYILDYFLFLNRYPEVGAHYVFVTLIGNFSEBAGMANICSSAG	924
Db	896	NKVIDIGLOQSLMPSNDSTVIANFSLAOYLHSGPPVFLVEEGNYSRKKQNMVCGMG	955
Qy	925	CNNEFSFTOKIOYATEFPEQSYLAIPIASSWDDPFIWILTP--SSCCRLYISGNKDKFCBST	983
Db	956	CDBNSLVQGIINAAELDTYTRVGRAPSSMIDYFDWMSPOSSCCRLY--NVTHQFNAS	1012
Qy	984	VNSINCLKNCKSIT-MGSRVPSVEQFHXYLFWLINDRNKICPYGGLANYSVNLISDG	1042
Db	1013	VMDPLCYR-CAPLIPBEGKQPOGKEFMKFLPMLFSDNPNPKCGKGHAAYGSAVNIAGDD	1071
Qy	1043	QVL-ASRPMATHKLKSQDTYTBALRAARELANITJADLRVPTGDPAFVEVPTITNVF	1100
Db	1072	TYIGATYEMTHHTIKTSADYTDAMKAKARLJASNITETMR-SKSD--YRVFYSVFEYF	1128
Qy	1102	YEOVLTILBEGFLMISLCTVPTPAVSCLLGLDLRSGLLNLTSLVMIILVDYVGFUALDI	1161
Db	1129	YEOVLTIIDTDFNLISVSLGISFLVTLVAVLCEGLMSAVIMCITITAMILVNNFGVNMIMG	1188

```

Oy      1162 SYMVASLNLVSAVGMVSVEFVSHTRSPALSTKPTWLEAREATISMGSAVPAGVAMTNL 12221
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1189 SLMAVSLNLVWMSCSISVEPFSCHTRATWTSTKGSRVASRAEALAHMGSSVFSGGITLTKF 1248

Oy      1222 PGIIVLGLAKAQILQIFFFRLNLILTLTGLLHGLVFLVILSYVSPDVNPA 1272
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1249 GGIIVLAPAKSQIFELFFYFRMYLNAVVLGATHGLIFLVLVLSYIGPSVNXA 1239

RESULT 4
US-09-462-136-6
; Sequence 6, Application US/09462136
; Patent No. 6426198
; GENERAL INFORMATION:
; APPLICANT: Carstea, et al.
; TITLE OF INVENTION: Genes for Niemann-Pick Type C Disease
; FILE REFERENCE: 4239-53894
; CURRENT APPLICATION NUMBER: US/09/462,136
; CURRENT FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: PCT/US98/13862
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/051,682
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1170
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-462-136-6

```

Query Match	19.2%	Score 1329	DB 2	Length 1170
Best Local Similarity	28.8%	Pred. No. 3.4e-116		
Matches 373	Conservative 230	Mismatches 512	Indels 182	Gaps 46
QY	10	LLMAL-----LLRLAQSEPYTHIQPGCAFEDECKNPF-----ELSGSLMTLSNVSCLSN	60	
DB	3	VLMIIIALVQGLMRLVQ-----TATCANVGNCGKKSVEGNELPCVPRSPSPRPVLS	54	
QY	61	TPARKITGDHLLLOKICPRLYTGNTQACCSAQQLVSLASLSITKALLTRCPACSDNF	120	
DB	55	ETSK-----LLVEYCGEEMKEVR-YACCTKQVVALRDNLQKAPLSSCPALCNF	105	
QY	121	VNLCHNTCSNGSLFINVTRVVAQGAQQLRAVAYAEAFQHSFQESYDSCSRRVRA	180	
DB	106	NNLFCHFTCAADQGRFVNITKY-EKSKEDKDI VAEILDVFNMSWMSSEYSDCKNIK	164	
QY	181	ATLAVGTWCYGSALCAOARMLNFGQDTGN--GLAPLIDTFHLEPQAAVSGIQPLNE	238	
DB	165	N-----GYAMDILGGAKKRYSGFLNKTGLDQAKMLGSSPQIINYKVDLANEE--KMGQEND	218	
QY	239	GVARCNSQGDVATCSCODCAASCPAIARPOALDSTFYLGQMPG--SLVLIITLGSVF	295	
DB	219	EVYACDDAQ-----YKCAQSDCOESCPHL-KP-LKDQGVCKVPLPCPSLSVLIFYTI	272	
QY	296	AVVITLLVGRFVARPAROKSKRVDPKKGSLSDK-----LSRSTH-----LLGQFPQMG	345	
DB	273	AFMYIYLCRRKKNKGMIVDDDIIVPESG-SLSESTNVESSNNENFENGKLANLFTKVG	331	
QY	346	TWVASWPLTIIIVLSYIPVALAAGLV-TELETTPDVELMASAPNSQAREKAFHOHFGPF	404	
DB	332	QFSVENPFIKILITTFYSIFVSFIIFQYATLETDPIINLMVSKNSEKFKREKYPFDNIGCF	391	
QY	405	FRTNQVILITABNRSSRYDSL--LLGPNFSGILDLDLLLELLELQERLRLHLYVWSPBAQ	462	
DB	392	YRTQDIFVNVETGPVLSYETLHMFDVENF-----ITTEL-----QSS	429	
QY	463	RNISLQDICVAPLNDNNTSLVDCINSLLQYFQNNRTLLLTATQYTMGQTSQYDMDYDHF	522	
DB	430	ENIGYDDICFRP-TEDSF-----CVTESFTQYFG-----ALPNKSKWREL	470	
QY	523	LYCANAPLTFPDGDTALSLACMADYCAPVFPFLAIGCYKGDYSEAEALIMFSLANNPAG	582	

[illegible]

```

RESULT 5
US-09-462-136-9
; Sequence 9, Application US/09462136
; Patent No. 6426198
; GENERAL INFORMATION:
; APPLICANT: Caretea, et al.
; TITLE OF INVENTION: Genes for Niemann-Pick Type C Disease
; FILE REFERENCE: 4239-53994
; CURRENT APPLICATION NUMBER: US/09/462,136
; CURRENT FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: PCT/US98/13862
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/051,682
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-462-136-9

```

Query Match	15.6%;	Score 1078;	DB 2;	Length 1296;
Best Local Similarity	24.0%;	Pred. No. 2.8e-92;		
Matches 321;	Conservative 237;	Mismatches 492;	Indels 288;	Gaps 41;

QY	76	ICPRLYGPMPNQAACCSAKOLVSLBASISITKALLTRCPASDNVNLHCNHTGS PNOL	135
Db	63	EFCPHLLTGDN-KLCTTSPQAEGLTKQJLAQAHHLIGRSCSCDNPAKLWCFTGSPNOD	121
QY	136	FINTVRVAQL--GAGOLPAVVAVEAF-----YQHS--PAEOSYDCSRVVRPAATLAV	186
Db	122	FVSISEMKEPIEKKEGFTEPEYQPAEVAVMTVEYRLSTDPABEGMFSSCKDVTGCGGALAV-	180
QY	187	TWCGYGGALCNAQWMLNFOGDTGNGL-APLDTIHLLEPGQAVSGGIQ- LINESVACN	244
Db	181	-MC--TSTPCTLTMWLEFYGTONLDLNIPIHTKFLVLDPIKTPSPRSDSTYMNVAFTGCD	236
QY	245	ESQGDVATVCSQDQDCAACPAIARPOALDSTFYLSOM-----PGLVLIILICSEAVVT	299
Db	237	KSARVGMFACSTSEC--NKEEYANLIDDDGKTQSQTQCNVHIGIACLNIFWMLAFGISLAV	294
QY	300	ILVVGFE-----RVAPARDKSNVDPKGTSLSDKLSFSTHTLLGGOF	341
Db	295	LLCVGFVFTSYDEYDTNLRQTOGSESPRNRK-----RTGAMI-----HNMEMNNA	342
QY	342	QGWGTWASWPLTILVLSVIPVVALAAGIVFELTTDPVELMSANSQARSKAPHDOHF	401
Db	343	RDIGMMAGRNPKSHFICAVLIFCLPMGIYHKESTNVNDWSSPRSRAROEWEVFNANF	402
QY	402	GPFPRTNQVILTPARSSRYDOSLLGPGNFGAILDLLELLELQSERLRLQWVSPEA	461
Db	403	GRPQRYQIOMLL--SHRDPQSGKLYCP-----VHKQIFELPITLAIKNISIQSDSG	455
QY	462	QRNLSLQDIQYAPLNPDNTSLYDCCINSILQYFQNNRTLLLTANQT-----L	509
Db	456	-RTILDDVCCRPMGPG---YDCLIMSPYVFOGNKEHLDMKSKKEFTVEBDDAPYF	510
QY	510	MGQISQVDPWKDHFLYCANAPLTFKQGTALASCMADYGAPVFPFLAIGYKCKDYSEBA	569
Db	511	SSEATTDMMNMMAACIDQPMSQK--TSGLSLSCMTYGGPSAPNM-VGQKSTYHNQAANS	567
QY	570	LIMFSLANNYPAGDRLAQAKLMEAPFEEMNAPFORMAGMQVFTERRSLDEINRTT	629
Db	568	IMTILVWQ--RTEBIOKALMEKEFLFKCEYERKEPKYV-BSFMAERSITTDIENDA	624
QY	630	ABDLPIFATSYVIFLYISLALGSY-----SSWSRWVDSKATILGJGVAVVLGAVMAM	684
Db	625	KDEIVTVALAFLIGYVTFSGRYFCENQMS-ILVHSR-----	664
QY	665	GPFSTLGRSSLVITQVVPFLVLSGADNIFFLVEYQRLRPBPREBEVHIGRLGVA	744
Db	665	GF-----	666
QY	745	PSMLTSLSEALCFPLGALTMPMAVTEFALTSGLAVIIDFLLOMSAFVALLSLDSKROBA	804
Db	667	-----TDLPAIRTECLYAGLAVLIDVWLCHTIFLALFPWMDQRELN	707
QY	805	SRLDVCCCVKQELPPPGQEGJL-----LLG-----FQKAVAPFLHMTTRGV	849
Db	708	G-----KP-BFFFPYQIKDILGAVYLIGRORATDFTWQGFHFOVAPFLMHMRTRIT	758
QY	850	LLFLFALFQVLSYMXCHTSVGLDQSLAPKQSYLLDPLFLNRYEYVGAQVYFVTTLGN	909
Db	759	GIIFLASFTTYVILSKISVGFQDSMAFLFEKYSITHTFYLDKFPDVGVPVFTVYDGLD	818
QY	910	FSSAEGMNAICSSACNNFSFQOKIQVATEPEQSGVILPSSWVDPIFIMLT--PSSCCR	968
Db	819	WHRPQVQKFCFFPCGSDTSFGNIMNVAVGHTEQYLTGSEMNMNIDNTYLEWTSRKSPPCK	878
QY	969	LYISGPNKDKPCPSFVNSL---NCLKNCMSITMGSVRBVSQFHLYLPWFLNDRPNIKC	1022
Db	879	VYVHDPN--TFQSTYRNKSLADKACRTCTMD-----	908

```

OY      PKGGLAAYSTSVNLTSJGCVLAASEFMAVYHPL--KNSODYTELRLARRELAANITDYLK 1082
Db      909 --DRAAFKODISTSRKIOASQFMTFHKKLSTSNSSDFIKADOTDRAWVSRRLERI-- 964
OY      1083 VPGTDPAAFEVPEPYTITTVFEYOYLTIPEGLFMISTCLVPTFAVSCLLJGLDRSGLLN 1142
Db      965 ----DPAHFAVAYSKIPEFEYOYSTIMPILTOLGFIIVGVEFGIICVTLGIDVKGAAV 1020
OY      1143 LSTVMILVDVYGFALMDISYNAVSLINVSASMSVEFVSHIRSRFAISGKPTWLERAK 1202
Db      1021 ICQVS-----NYHIVSS-GILFEFSVNVUKGYACISRQAKRABE 1060
OY      1203 EATISMGSAVAGVAMTNLPGELIVLGLAKAQLOIEFFRLNLTLLGLLGLVFLPVIL 1262
Db      1061 STVSGISGIIILSGVVTVMAGSTMFLSGAHQOIIIVVEFFKFLITIVSSAVHALITIELL 1120
OY      1263 SYVGPDVNPMALALEQKRAEVAALVAAMVASCNHPRSRSTADNIYVN-----HSEFSGS 1314
Db      1121 AFQGSRGHSGSETSTNDNDEQHDACVLS--PTAESHSNVEBGLINRPSLLDASHIIDPL 1178
OY      1315 IKGAAGAISNELLNNGROP 1332
Db      1179 LKAEGGIDKAI---GRDF 1193

```

RESULT 6  
US-08-540-406-19

```

1 Sequence 19, Application US/08540406
2 Patent No. 5837538
3 GENERAL INFORMATION:
4 APPLICANT: SCOTT, MATTHEW P
5 APPLICANT: GOODRICH, LISA V
6 APPLICANT: JOHNSON, RONALD L
7 TITLE OF INVENTION: Patched Genes and their Use
8 NUMBER OF SEQUENCES: 19
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Fleht, Hobbach, Test, Albritton & H
11 STREET: Four Embarcadero Center, Suite 3400
12 CITY: San Francisco
13 STATE: CA
14 COUNTRY: US
15 ZIP: 94111
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: PatentIn Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/540,406
23 FILING DATE: 06-OCT-1995
24 CLASSIFICATION: 435
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Rowland, Bertram I
27 REGISTRATION NUMBER: 20015
28 REFERENCE/DOCKET NUMBER: a60190-1
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 415-781-1989
31 TELEFAX: 415-398-3249
32 INFORMATION FOR SEQ ID NO: 19:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 1447 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 MOLECULE TYPE: 'protein
39 US-08-540-406-19

```

Query Match 9.4%; Score 651.5; DB 1; Length 1447;  
 Best Local Similarity 22.0%; Pred. No. 1.le-51;  
 Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
 225 PGAVAGSIQIPLNRCVAFNCESQGDVYATCSQDCAASCPIAIPALDSTFYIGQMPGS 284

```

Db      24 PCRPAAGGRRRTGCGRRRAAPDRD-----YLRBSYCDAAFALEOI---- 65
Qy      285 LVLIILLCVFAVVITLLVGFRAVARADSKMVDPK-----KGTSLSDKLSFSTHTLL 337
Db      66 -----SKGKATGRKAPLMRLAKPKQRLFLKLGCIYQKNC 98
Qy      338 GGFPGCGWGWASWPLTILVLSVTPVVALAAGLVFTELTDDPVELMSAPNSQARSKAH 397
Db      99 GRF-----LVVGLILFGAFAVGIKKAANLETNVEELWVEVGVRSLRYNT 143
Qy      398 DGHFGFPRFTN-QVILTAPNRSSYRDSLLGKNSGLIDLLLELE--LQERLRL 454
Db      144 RQKIGSEAMFNQOLMTQPKESG-----ANVLTTEALQHLDSALQASRAV 190
Qy      455 QVWSPBAQNNISLQICYAP-----LNPNTSLYDCINSLSLQVFQNNRTLLTLTA 505
Db      191 YWYN-----RQWKLHHCYKSGELITETGVMDQILEVLPCLITPLDCFWBGAKLQSGTA 246
Qy      506 NOTLMGQ-----TGQVMDKHFLY-----CAN----- 527
Db      247 --YLIGKPRPLRWTFPDLFLEBELKKINYQVDSWEMELNKAEGVGYMDRPTCLNPADPDC 304
Qy      528 -----APL-----TFPKGT-----ALALSCMA 544
Db      305 PATAPNKSTKPLDUALVLANGCGHLSRKYMKQEBELVYGIVKSTGKLVSAHALQTNF 364
Qy      545 DYGAVPFPLAIGGYKDKDYSEAEALIMTFSLNNYPADGPRLAQAKLMEAPLEEMRAQ 604
Db      365 QLMTPRKQW--EHFQGEY-----VSHNNME--DKAAAILLEAQRTYVEVHQSV 411
Qy      605 RRMAGHFQVFTFAERLEDEINFTTAEDELPIFATSTIVFLYISLALSGYSSSRWMDV 664
Db      412 AONSTOKVLSFTT-TLTDILKSPSDSVYIRVASGLMLMAVACLTM--LRMD--CSKS 465
Qy      665 KATLGLGVAVVLGAAVMAAGFSPSYGISLVLTVPPFLVSGADNIFIVLEYQRL 724
Db      466 QGAVGLAGLVLAALVAAGLGLCSLIGISFNAAITQVLPALAGVVDVDPALAAHAFST 525
Qy      725 PRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPMAVTFALTSGLAVILDF 784
Db      526 GQNKRIPEDRDRGECLKTGASVALGISNVTAFFMAAILIPALAAFLQAAVVVVFNF 585
Qy      785 LLOMSAFVALLSLDSKQESRLDVCC-----VKRQEL-----PPRG 822
Db      586 AAVLLIFPALISMDLYRDRDLIDICFTSPCVSRVIOVEPQAYTDTHTNTRYSPBPY 645
Qy      823 QGEGE----- 827
Db      646 SHSFAHETQITMQSTVOLRTEYDPTHVYTTAERSEISVQPVTVTODTLSCQSPST 705
Qy      828 -----LIGFPOKAVAPFLHMTITRGVLLFLALFGVSL 861
Db      706 SSTRLDLSQFSDSLHCLBPPCTKMTLSSPAEHGVAPFLKKPRAKVVIFFLGLGVL 765
Qy      862 YSKCHISVGLDOELALPKDSYLLDYFLFNRYFEVGAQPVYFTTTLGYNFSSBAQNAICS 921
Db      766 YGTRVRDLDLDLDIVPRETREYDFYNAQFKYFSF-----YNN----- 803
Qy      922 SAGCNNSFTOKIQVAT-----EPPEOSYLAIPASS-----WVDDFIDWL----- 961
Db      804 -----YIVYQKADYFNIOHLHYDLHRSFNSVAKTVLAEKKOLPKKMLHYFDMDLGLOD 857
Qy      962 -----TPSS-----CCRLYISGPNKDKFCPTVNSLNCIAKNCMSITMGS 1000
Db      858 AFDSDMETGKIMPNVNNKINGSDGCVLAKLVNCGSDK--PIDISGLT--KQRLVDADI 913
Qy      1001 VRPSVEQPHKYL-PWFLND-----RPNIKCKPKGGLAY--STSVNLTSDDQ 1043
Db      914 INPSA--FYIYLTAWNSNDPVVAAQANIRPHRPBWVADK--ADYMEETRLRIPAAEP 968
Qy      1044 VLASRMAVHKPLKJQSDYTEALRAABELAANITA-DLAKPQTDPAFVFPYTTINVY 1102
Db      969 IEYADPFYLANGRLDTSDEVALEKVRTICSNYTSIGLSSYPNG-----YPP---LFW 1018

```

```

Qy      1103 EGYTLTBEGLEFMLSCLVPTFAVSCLLGLDLRSGLNLNLSIWNILVDTVGFMAIWDIS 1162
Db      1019 EGYIGLRHMLLFIISVLACFLVCAVFLNLPWTAGII-VWVIALMTVELFGMGLIGIK 1077
Qy      1163 YNAVSLINLUSVAGSVFVSHITSF--AISTRTMLEBAKEATISGSAVFAVANTN 1220
Db      1078 LSAPVVILIASVGIQVEFTVVAALAFYLAIGDKN--RRAVLALEHMFAPVLDG-AVST 1133
Qy      1221 LFGIIVLGLAKAQLQIFFFRNLNLTLLGLHLGVLPVLISYVG--PDVNPALALBO- 1277
Db      1134 LIGVLMLAGSEFDFVRYFFAVLAILTLGLVNLGLVLLPVLLSFGRPPRPSPANGRL 1193
Qy      1278 --KRAEAVAAVMAVASCNHPBSRVSTADNIYVNSFECSIKG 1317
Db      1194 PTPSPPEPPSVVRFAMPQHTH--SGSDSDSEYSSQTTVSG 1233

```

## RESULT 7

US-08-656-055-19

Sequence 19, Application US/08656055

Patent No. 6027882

GENERAL INFORMATION:

APPLICANT: SCOTT, MATHEW P

APPLICANT: GOODRICH, LISA V

APPLICANT: JOHNSON, RONALD L

TITLE OF INVENTION: Patched Genes and their Use

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSER: Flehr, Hohnbach, Test, Albritton &amp; Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: CA

COUNTRY: US

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/656,055

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/540,406

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Rowland, Bertlem I

REGISTRATION NUMBER: 20015

REFERENCE/DOCKET NUMBER: A60190-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989

TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 1447 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-656-055-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;

Best Local Similarity 22.0%; Pred. No. 1,1e-51; Indels 421; Gaps 47;

Matches 300; Conservative 181; Mismatches 460;

```

Qy      225 PGQAVSGGIQPLNCGVARNCSQGDVATCSQDCAASCRAIARQALDSTFYLGQMPGS 284
Db      24 PCRPAAGGRRRTGGLRRAAAPDRD-----YLRBSYCDAAFALEOI--- 65
Qy      285 LVLIILLCVFAVVITLLVGFRAVARADSKMVDPK-----KGTSLSDKLSFSTHTLL 337

```



[illegible]

```

QY      1163 YNAVSLINLVASNGMSVEFVSHIRSF--AISTPTMLEPKKATIISMGSVFPAGVAMTN 1220
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::|
Db      1078 LSAAPVFILLASVGIGAEFFTHVALAFALTALGDKN---RRVIALLEHMFAPVLDG-AVST 1133

QY      1221 LPGLIVLGLAAQAQIQIUFFPRNLNLTLLGLHLGIVEPLVISYG--PDVNPAALAEQ- 1277
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::|
Db      1134 LLGVILMAGSEFDIVRYFFFAVLAIIITLIGLNBLVLLPVLSIFGPYPPEVSPIANGNRL 1193

QY      1278 --KRAEVAAVWVASCPNHPSRVSTADNIYNHSFEQSIXG 1317
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::|
Db      1194 PTPSEPPEPPSVVRPAMPFGHTH--SGSDSSDSBSYSOTVSG 1233


RESULT 8
US-08-954-668-19
; Sequence 19, Application US/08954668
; Patent No. 6172200
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley, Hoag & Eliot
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,668
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36709
; REFERENCE/DOCKET NUMBER: SUV-003.06
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-954-668-19


Query Match          9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. No. 1.1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

QY      225 PGQAVGSGTIGPLNGVACRNESGGDDVATCSGCDCAASCAPAIPARQALDSTFYLGMPGS 284
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::|
Db      24  PGRARGGRRRTTGTLRAAAPRD-----YLHRPSVCDAFALEGI---- 65

QY      285 LVLIILICVPAVVTTILVGRVAPARDKSMVDPK-----KGTSLSDKLSFTSTTL 337
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::|
Db      66  -----SKGATGRKAPLMRARFORLLFLGCGYIOKNC 98

QY      338 GQFPGMGWTWASWPLTIIVLSVIPVALLAAGLVETELTDPVELWSAPNSQASEKAFF 397
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::|
Db      99  GRF-----LVVGLIIFGAFAVGLKAAIDLETVEELMWVGVGRVSRRELANTY 143

QY      398 DQHGFPPFRFN-QVILTAPNNSSVRYDSSLLAGPNFSGIIDLDLLELLE--IQERLRHL 454

```



```
Db      144 RQKIGBEAMFNQMLQTPKEEG-----ANVLTEALLQHLDSALQASRVHV 190
Qy      455 QVMSPEAQRNLSLQDICYAP-----LNPNTSLYDCCINSLSLQYFQNNRTLLLLTA 505
Db      191 YMYN-----RQWLEHLQYKSGELITETGYMDQILEVLPCLITPDLDCFWGAKLQSGTA 246
Qy      506 NOTLMGQ-----TSQVDMKDFLY-----CAN----- 527
Db      247 --YLLGKPLRWNTNPDLEBLELKKINYQVDSWEMLNKAEGVGYMDRPCINPADPDC 304
Qy      528 -----APL-----TFKQGT-----ALALSQMA 544
Db      305 PATAPNKSTKPLDMLALVINGGCHGLSRKYMHOEELIVGTVKNSGTGLVSHALQTMF 364
Qy      545 DYGAVPFPPLAIGYKGDYSEAEALIMTFSLNNYPADGPRLAQAKLMEAEFLEEMRAFO 604
Db      365 QLMTPKQWY---EHRFGYEV-----VSHINWNE---DKAAALIEAMQRTYVEVHOSV 411
Qy      605 RRMAGNFQVTPFAERSLEDEINRTAEDLPFATSYIVIFLYISLALGSSYSMSRVWVS 664
Db      412 AONSTQKVLSTFT--TLLDILKSPDVSIVKVASGYLLMALVACLTM---LRMD--CSKS 465
Qy      665 KATLIGGVAVVLGAAVMAAGFFSYLGIRSSLVILQVPELVLSGADNIFIFVLEYORL 724
Db      466 QGAVGLAGVLLVALVAAAGLGLCSLIGISFNAATTQVLPPLAGVGDVDFLLAAHAFSRT 525
Qy      725 PRRPGEPRVHIGRALGRVAPSMILCSLSEALCFEFLGALTMPMPAVTFTGLAVILDF 784
Db      526 GQNKRIPEFDRGECIKRGTASVALISINVTAFPMALILPILALRAFLLQAAVVVVFNF 585
Qy      785 LLQMSAFVALLSDSKQBPASRLDVCC-----VKQDEL-----PPPG 822
Db      586 ANVLIFPAILSMIDLRYRREDRLRIDPCETSPCVSHVIOVBQAYVTDHNTNRYSPBPY 645
Qy      823 QGEBL----- 827
Db      646 SSHSFAHETQITMQSTVQLRTEYDPTHVYTTAEBRSSISVQPVTTQDTLSCQSPBST 705
Qy      828 -----LIGFQKAYAPFLHMTTRGVLLFLFALFGVSL 861
Db      706 SSTRLDLSQPSDSLHCLEPCTKMTLSFPAEKHYAPFLPKPKAKVYVFLFLGLGVL 765
Qy      862 YSMCHISVGLDOELALPKDSYLDYFLPLNRYFEVGAPVYFVTTGLGYNSSSEAGNAICS 921
Db      766 YCTTRVRDGLDLDIYPRETREVDFIAQFKYFSF-----YNN----- 803
Qy      922 SAGCNPFSTQKIQVAT-----EFPEOSYLAIPASS-----WVDFETDML----- 961
Db      804 -----YIYTOKADYENIQHLHYDLHRSFSNVKYVMLEBNKOLPKXWMLHYFRDMLQGLQD 857
Qy      962 -----TPSS-----CCRLYISGPNKDKFCPTVNSLNLCKNMSITWMS 1000
Db      858 AFDSDMENGKIMPNYKNKSGDDGVLAAYKLIVQSGDK--PIDISGLT--KQRLVYADGI 913
Qy      1001 VAPSEVQPHKYL--PWFELND-----RPNIKCPKGGLAAY--STSVNLTSDQ 1043
Db      914 INPSA--FYIYLTAWNSNDPVAAVAAQANIRPHRPBWVHDK--ADYMPETRLRIPAAEP 968
Qy      1044 VLASRPMAHKPLKNSQDYTEALRAARELANITA-DLAKVPGTDPAFEVFPRTINNVY 1102
Db      969 IEVAPGPFPLNGLRDTSDFEALIEKVRTICSNYTSIGLSSYPNG-----YPP---LFW 1018
Qy      1103 EGYLTLPFGLFPMLSCLVPTFAVSCLLGLDLRSLGMLNLSTVMILVTVGFMALMDS 1162
Db      1019 EGYIGRHHLLLFISVYLACTPLVCAVFLINPWTAGIIT--VMYALAMTVELFGMGLIGIK 1077
Qy      1163 YNAVSLINLSVAGKSVSEFVSHITRSF--AISTKPTMLERAKEATISMSAVFAGVAMTN 1220
Db      1078 LSAVPLVILIASGIGVEFTVHVALAFLTAIGDKN---BRAVALLEHMPAPVLDG--AVST 1133
Qy      1221 LGGILVYGLAKAQOLIGIFFPRNLNLTLTGLLHGLVFLPVILSYVG--PDVNPALALBQ- 1277
```

```
Db      1134 LIGVMTLAGSEFDFIVRFPAVLAITLTLGVNLGVLLPVLLSFEQPYEVSFANGLNRL 1193
Qy      1278 --KRAEEAVALVAAVMAVASCNHPBSRVSTADNIYVNSFEQSIGK 1317
Db      1194 PTPSEPPPSVYRFAMPQGHTH--SGSDSPSEVSSQTTVSG 1233

RESULT 9
US-09-268-140-5
; Sequence 5, Application US/09268140
; Patent No. 6268176
; GENERAL INFORMATION:
; APPLICANT: Gemmili, Robert M.
; APPLICANT: Diadkin, Harry A.
; TITLE OF INVENTION: TRC8, A GENE RELATED TO THE HEDGEHOG RECEPTOR, PATCHED
; FILE REFERENCE: 93445-00004
; CURRENT APPLICATION NUMBER: US/09/268,140
; PRIOR FILING DATE: 2000-03-12
; PRIOR APPLICATION NUMBER: US 60/077,723
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1447
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-140-5

Query Match      9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. No. 1,1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

Qy      225 PGQAVSGGIQPLNEGVARCNESQGDVATCSCQDCASCPAIPQALDSTFYLGQMPGS 284
Db      24  PGRPAGGRRRRRTGGLRBAAPDRD-----YLHRPSYCDAAFALEQI--- 65
Qy      285 LVLIILISVPAVVTIILVGRVAVAPAKSKKNVDPK-----KQTSLSDLSPSTHLL 337
Db      66 -----SKGRATGRKAPLMRAKFORLLFKLCYIQKNC 98
Qy      338 GQFQGWGTWVAWPDLTILVSVIPVALAAGLVTELTTPDVELMSAPNSQARESEKAFH 397
Db      99 GKF-----LVGGLIFGAFAVGLKAALETNVEELMVEGVRSBELNYT 143
Qy      398 DQHFPPFRITN-QVILTAPNRSSYRYDLSLGPKNFSGILDLDLLELLE--LQERLRL 454
Db      144 RQKIGBEAMFNQMLQTPKEEG-----ANVLTEALLQHLDSALQASRVHV 190
Qy      455 QVMSPEAQRNLSLQDICYAP-----LNPNTSLYDCCINSLSLQYFQNNRTLLLLTA 505
Db      191 YMYN-----RQWLEHLQYKSGELITETGYMDQILEVLPCLITPDLDCFWGAKLQSGTA 246
Qy      506 NOTLMGQ-----TSQVDMKDFLY-----CAN----- 527
Db      247 --YLLGKPLRWNTNPDLEBLELKKINYQVDSWEMLNKAEGVGYMDRPCINPADPDC 304
Qy      528 -----APL-----TFKQGT-----ALALSQMA 544
Db      305 PATAPNKSTKPLDMLALVINGGCHGLSRKYMHOEELIVGTVKNSGTGLVSHALQTMF 364
Qy      545 DYGAVPFPPLAIGYKGDYSEAEALIMTFSLNNYPADGPRLAQAKLMEAEFLEEMRAFO 604
Db      365 QLMTPKQWY---EHRFGYEV-----VSHINWNE---DKAAALIEAMQRTYVEVHOSV 411
Qy      605 RRMAGNFQVTPFAERSLEDEINRTAEDLPFATSYIVIFLYISLALGSSYSMSRVWVS 664
Db      412 AONSTQKVLSTFT--TLLDILKSPDVSIVKVASGYLLMALVACLTM---LRMD--CSKS 465
Qy      665 KATLIGGVAVVLGAAVMAAGFFSYLGIRSSLVILQVPELVLSGADNIFIFVLEYORL 724
Db      466 QGAVGLAGVLLVALVAAAGLGLCSLIGISFNAATTQVLPPLAGVGDVDFLLAAHAFSRT 525
Qy      725 PRRPGEPRVHIGRALGRVAPSMILCSLSEALCFEFLGALTMPMPAVTFTGLAVILDF 784
```

526 GQKRIPEDRGECKRTGASVALTSISVTAFFMAALPIPALAFSLQAAVVVVF 585  
785 LLOMSAFVALLSDSKROEASRLDVCC-----VKPOL-----PPG 822  
586 ANVLIFPALISMDLYRREDRLDIFCCFTSPCVSRVIOVEPQAYDTHDNTRYSPPEY 645  
823 QOEGE----- 827  
646 SSSHFAHETQITWQSTVQLRTEYDPHTHYTTAABRSEISVQPTVTOPTLSCQSPST 705  
828 -----LGFQKAYAPFLHMTTRGVLLPLALGVSL 861  
706 SSTRDLLSQPSDSLHCLBPPCTKWTLSSEAEKGYAPFLKPRAKVAVVFLFGGLGVSL 765  
862 YSMCHISVGDQELAPKDSYLLDYFLFLNRYPEVQAPVYVTLGNFSSSEAGNNAICS 921  
766 YGTRRDGLDLTDVPRRETRVDFTAAQKTFSS-----YNN----- 803  
922 SAGCNNSFTOKIQVAT-----EPBQSYLAIPASS-----WVDDFIDWL----- 961  
804 -----YIVTQKADYDNIGHLVDLHRSFENVKYMLEBKQLPKWMLHYFRDMLQGLQD 857  
962 -----TPSS-----CCRLYISCPNNDKFCPSTVNSLNCCKMCSITMGS 1000  
858 AFDSMDETKIMNNYKNGSDGVLAVKLVQGSBDK--PIDISQLT--KORLVADAGI 913  
1001 VRPSVQFHKYL-PWFLND-----RPNKCPKGLAAY--STSVNLTSDDQ 1043  
914 INPSA--FYIYLAWSNDPVAAASOANRPHREPWMDK--ALYMETRIRKIPAAEP 968  
1044 VLASRFMAVHKPLKNSQDYTEALRAARELANITA-DLRKVPGTDPAFVFPYITINVPY 1102  
969 IEVAGPPFLNGLRDTSDFEALTEKVRTICSNVTSIGLSYPNG-----YFP-----LFW 1018  
1103 EGYLTTLPEGLFMLSCLVPTPAVSCILGLDRLSGILMLSLVMLVTDVGMALMDIS 1162  
1019 EGYIGLRHMLLFISVLACTFLVCAVFLNPMPTAGII--VMVLAMLTVELFEGMGGLIGIK 1077  
1163 YNANVSLINVSAAVMSVEFVSHITRSP--AISTKPTMLBRKKEATI SMGSAVPAVGMATN 1220  
1078 LSHAPVAILLASVIGIEFTVHALAFLNIGDKN--RRAVIALHEMFAPVLDG--AVST 1133  
1221 LSGILVGLAKAQLOIIOFFERLNLITLGLHLGLVFLPYILSYVG--PDVNPALAEQ- 1277  
1134 LSGVLMLASGEPFIYRIFPAVALIITLIGVNLGLVLPVLPSFPGPYEVSANGLNLN 1193  
1278 --KRAEAAVAAMVASCPNHSKVSADNITYNHSFEGSIKG 1317  
1194 PTPSPPPSVVVFAMPFGHTH--SGSDSDSESYSSQTVVSG 1233

RESULT 10  
US-08-918-658-19  
; Sequence 19, Application US/08918658  
; Patent No. 6429354  
; GENERAL INFORMATION:  
; APPLICANT: SCOTT, MATHEW P  
; GOODRICH, LISA V  
; JOHNSON, RONALD L  
; TITLE OF INVENTION: Patched Genes and their Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fiehr, Hobbach, Teet, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/918,658  
FILING DATE: 22-Aug-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/656,055  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/540,406  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertram I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: a60190-1  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1447 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-08-918-658-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1,1e-51;  
Matches 300; Conservative 101; Mismatches 460; Indels 421; Gaps 47;

225 PQAVSGGIQPIINEGVARNESQGDVATSCQDCAASCRAIARPOLDSTFYLGQMPGS 284  
24 PQBPAGGGRRRRTGRLRAAPDRD-----YLHRSYCDAAAFABQI--- 65  
285 LVLIILICVPAVVTLLVGFVAPARDKSKAVDER-----KGSLSDKLSFSTHTLL 337  
66 -----SKKATGRAPMLRAKQRLFKGCVYIQKNC 98  
338 GQFQMGTVASWPLTILVIVPVALAGVFTLTTPDELMSANSOAREKAP 397  
99 GKF-----LVGGLIFGAPAVGLKAALETNVEBELWVEGVRSRLANT 143  
398 DQHFQFPFTN-QVILTAENRSSYRYSLLGPKNFSGLIDLLLELLE--LQERLRH 454  
144 RQKIGSEAMFNQMLQTKERG-----ANVLTTEALLQHLDSALQASRVHY 190  
455 QVWSPBAQNISLODICVAP-----LNPNTSLYDCCINSLSLQYQNNRTLLTLTA 505  
191 YWYN-----RQWLEHLCHYSGELITETGYMDQIIEYLPCLITTPIDCFWBGAKQSGTA 246  
506 NOTLMGQ-----TSQVWMDHFLY-----CAN----- 527  
247 --YLLGKPLRKMTNPDLEFLBELKKINYQVDSWEMELKABVGHGYMDRPLCLNAPDPDC 304  
528 -----APL-----TFKQGT-----ALALSQNA 544  
305 PATAPRNKSTKPLDMALVLNGCHGHSRKVMQEBELIVGVYKNSTGLVSHALQTFN 364  
545 DYGAVPFPLAIGGYKQDYSEABALIMTFSUNTPAGPRLAQAQKMEAPLEBRAQ 604  
365 QLMTPKQMY--EHFGYGEY-----VSHINNE--DKAAALIEAMORTIVEVHQSV 411  
605 RMAMGQVQYTPAESLBEIRRTAEDLPFATSYIVIFLYISLALSGSYSSSRVWVS 664  
412 AQNSTOKVLSFTT--TLDDILKSPDSVSYIRVAGVLLMLAACLTM---LKNQ--CSGS 465  
665 KATLIGGVAVVLGAVMAAMGFFSYLGRSSVLVIVVPEFLVSYGADNIPFLVLEYORL 724  
466 QGAVGLAGVLVLAASVAGLGLCSLIGISFNATQVLPFLAGVVDVFLAHAFSST 525  
725 PRRGEPREVVHIGRALGRVAPSMILCSLSEALCFPLGALTTPMPAYRTPLTGLAVILDF 784

Db 526 GGNKRIPEEDRTGECCLKRTGASVALTISINVTAFPMALIPIPALRAESLOAAVVVNF 585  
Qy 785 LLQMSAFVALLSLDSKQASRLDVCC-----VKQDEL-----PPRG 822  
Db 586 ANVLIFPAIISMDLYRREDRLDIFCCFSPCVSRVIVVEQAVYDTHNTRYSPPEPY 645  
Qy 823 QGEBL----- 827  
Db 646 SSHAHEQTITMGTQVQLRTEVDPHTHYTTAERSEISVQPVYVTOTDLSQSPEST 705  
Qy 828 -----LLGFQKAVAPFLLHMTTRGVVLLFLALGVSL 861  
Db 706 SSTRLDLSPPSSSLHCLPEPCCKTLLSFAEKHYAPFLIKRAKVVYFLFGLGVSL 765  
Qy 862 YSMCHISVGLDQELAKDSYLLDYELFLNRYFEVQAPYFVYTLGYNSSSEAGNNAICS 921  
Db 766 YGTRVRDGLDLTDLYPRTREYDFIAQPKYFSF-----YNN----- 803  
Qy 922 SAGCNNSFTQKIQYAT-----EPPEOSYLAIPASS-----WVDFIDWL----- 961  
Db 804 -----YITQKADYFNIGHLLYDLHRSFSNVYVWLEENKOLPKWMLHYFRDWLQGLQD 857  
Qy 962 -----TPSS-----CCRLYISGPNKDKCPSTVNSLNLCKXCMSTWGS 1000  
Db 858 AEDSDMETGKIMPNKNGSDGCVLAKVLVOTGSHDK--PIDISQLT--KORLVADAGI 913  
Qy 1001 VAPSVGEQFHLYL-PWEFLND-----RPNIKCPKGGLAAY--STSVNLTSDDQ 1043  
Db 914 INPSA--FYIYLLAWNSNDPVAAASQANIRHPRPFWMDK--ADYMETLIRIPAAEP 968  
Qy 1044 VLASRFMAVHKPLKOSDYTEALRAARELANITA-DLARKPGTDBAEVFPYTTINVY 1102  
Db 969 IEYAFPFPLNGLRDTSDEVEAIEKVRTICSNVTSGLSSYPNG-----YPF-----LFW 1018  
Qy 1103 EGYLTLPBGLFMLSCLVPTFAVSCLLGLDLSGLMLSLVIMLVYDVGPMALMDS 1162  
Db 1019 EGYIGIRHMLLFISVLAICTFLVCVFLINPWTAGIT--VNVLAITVTLFQMGGLIGIK 1077  
Qy 1163 YNAVSLINLVSAVGSVEFVSHITRSF--AISTKPTMLERAKEATISMSAFAVAGVAMTN 1220  
Db 1078 LSAVPVILIASGVGEFTVHALAFLAIGDKN--RAVLALEHMPAPVLDD-AVST 1133  
Qy 1221 LGGILVLAGAKQLOIIFFRILNLITLGLHGLVFLVYISYVC--PDVNPALALEO- 1277  
Db 1134 ILGVMLAGSEBDFIYRYFFAVLAITLIIIGVNLGLVLPVLLSFEGPYEVSFPAQLNRL 1193  
Qy 1278 --KRAEAAVAAMVAVSCPNHPSKSVSTADNIIYVNHSESGSIRK 1317  
Db 1194 PTPSPPEPPSVVRFAMPQGH--SGSDSSDSSEYSQTTVSG 1233

RESULT 11  
US-09-724-631-19  
Sequence 19, Application US/09724631  
Patent No. 6551782  
GENERAL INFORMATION:  
APPLICANT: SCOTT, MATHEW P  
GOODRICH, LISA V  
JOHNSON, RONALD L  
TITLE OF INVENTION: Patched Genes and their Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fleury, Hobdach, Teat, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/724,631  
;; FILING DATE: 28-Nov. 6551782-2000  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/656,055  
;; FILING DATE: 1996-05-31  
;; APPLICATION NUMBER: 08/540,406  
;; FILING DATE: <Unknown>  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Rowland, Bertam I  
;; REGISTRATION NUMBER: 20015  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-781-1989  
;; TELEFAX: 415-398-3249  
;; INFORMATION FOR SEQ ID NO: 19:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1447 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-724-631-19  
  
Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. NO. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
  
Qy 225 PQGAVSGGIQPLNBNVARNCSQGDVATCCQDCACAPAIARQALDSFTYLGOMGS 284  
Db 24 PGRPAGGRRRRRTGIRBAAPDRD-----YHRSYCDAAFALEQI--- 65  
Qy 285 LVLIILISVPAVVITLVGFPAVARADSKKAVDPK-----KGSISDLKSFSTHTLL 337  
Db 66 -----SKGATGKAPLMRLAKQRLFLKIGCTIQKNC 98  
Qy 338 GQFQGWGTWVAWDLTILVLSIPVVALAAGLVETELTDPVELMSAPNSQARSEKAFH 397  
Db 99 GKFE-----LVVGLLIFGAFVAGLKAANLETNVEBELWVGSRVSRRLNYT 143  
Qy 398 DQHFPPFRFTN-QVLTLPNNSSTRYDSLLGPKNFSGILDLDLLELLE--LQERLRL 454  
Db 144 RQKIGEEAMFPQMLIQPKKEG-----ANVLTTEALLQHLDSALQASRVHV 190  
Qy 455 QVWSPDEARNISLIDICYAP-----LNPDNSTLYDCCINSLLQYFQNNRTLLLT 505  
Db 191 YKYN-----RQMKLEHLCYKSGELITETGYMOQIIEYLPCLITPDLDFWEGAKIQSGTA 246  
Qy 506 NOTLMGO-----TSQVDWKDHPLY-----CAN----- 527  
Db 247 --YLLGKRPMLKMTNDPLFEBELKKIYQYDSWEMLNKAEVGHGMDRCLNPADPDC 304  
Qy 528 -----APL-----TFKDG-----ALALSCMA 544  
Db 305 PATAFNKSTKPLDMALVINGGCHGLSRKYNHMOBELIVGGIVKNSSTKLVSANHLQTMF 364  
Qy 545 DYCAVVPFPLAIGVKYDSEABALIMTFSLNNVPADPPLAQAUKMEAFLEEMRAFQ 604  
Db 365 QLMTRKQNY--EHKGYEY-----VSHINWNE--DKAAALIEAMQRTVEVHOSV 411  
Qy 605 RRMAGMFOVTEAERSLDEINRTAEDLPFATSYIVIPYISIALGSSWSRWVYDS 664  
Db 412 AQNSIQKVLSPFTT--TTLDILKSPSDVIVIVASGYLLMLVACTM--LRMD--CGKS 465  
Qy 665 KATLGLGVAVVLGAVMAAGFPSTYLGIRSSLVLIQVVPFLVLSYGADNIFFLVEYQRL 724  
Db 466 QGAVGLAGVLVLAISVAGLGLGCSLIGISFNAATQVVPFLALGVDVDFLAAHAFST 525  
Qy 725 PRRPEPEPEVHIGRALGVAPSMILCSISEAICFELGLTMPAPRTALNSGLAVIIDF 784  
Db 526 GGNKRIPEEDRTGECCLKRTGASVALTISINVTAFPMALIPIPALRAESLOAAVVVNF 585



Db 646 SSHAHEQTITMSTVOLRTEYDHTHYTTTAEPRSEISVQPTVTOTDLSQSPBST 705  
Qy 828 -----LIGFQKAVAPFLHWITRGVLLFLALFGVSL 861  
Db SSTRDLDSQFSDSLHCLERPCWKMTLSSFAEKHAPFLKPKAKVVIFFLGLGVS 765  
Qy 862 YSMCHISVGLDDELALPKDSYLLDFLEFLNRYFEVGAVPYVTTLGYNSSSAGNAICS 921  
Db YGTTVRDGLDLDIVPRETREVDFIAQFKYFSF-----YNN----- 803  
Qy 922 SAGCNNSFTOKIOYAT-----EPFQSYLAIPASS-----WVDFIDWL----- 961  
Db 804 -----YVTQKADYPNIOHLLYDLHRSFNKTYMLEENKOLPKMWLHYFRDMLOGLOD 857  
Qy 962 -----TPSS-----CRLYISGPNKDKFCSTVNSLNCNKMSITWGS 1000  
Db 858 AFDSMETGKIMPNNYKNGSDGVLAAYKLLVOTGSRDK--PIDISQLT--KQRLVDADGI 913  
Qy 1001 VAPSVQGFHKYL--PWFLND-----RPNIKCPKGGLAAY--STSVNLTSDGQ 1043  
Db 914 INPSA--FYIYLTAVNSNDPVAYAAASQANIRPHRPEWHDK--ADYMBETRLRIPAAR 968  
Qy 1044 VLASRFMAVHKPLKNSODYTEALRAARELANITA--DLRKVPOTDPAFEVPPYTTINVFY 1102  
Db 969 IEYAQFPFLNGLRDTSDEVEALEKVRTICSNYTSIGLSSYPNG-----YFP---LFW 1018  
Qy 1103 EGYLTLPBGLFMLSCLVPTFAVSCLLGLDLRSGLLNLISIMILVDVTFGMLMDIS 1162  
Db 1019 EGYGLRHWLLLFISVVLACTFLVCVAFPLNPMPTAGII--VMVLALMTVELFGMGLIGIK 1077  
Qy 1163 YNAVSLINVSAGMSVEFVSHITRSF--AISTKPTWLERAKEATISGSAVPAVAMTN 1220  
Db 1078 LSAAVPAVLIASVGIQVEFTVHVALAFLPAIDKN---RRAVLALEHMFAPVLGD--AVST 1133  
Qy 1221 LFGIIVLGLAKQOLQIIFPRLNLITLGLHGLFVLISYVYG--EDVNPALALBEO- 1277  
Db 1134 LFGVLMAGSEDFIVRYFAVLAITITIGVNLGLVLLPVLLSFPQRYEVSFPAAGLNR 1193  
Qy 1278 --KRAEAVAAVMVASCPNHPRSVSTADNIYVNHSPFSGIKG 1317  
Db 1194 PTPSPPEPPSVVRPAMPPEGHTH--SGSDSDSEYSQGTIVSG 1233  
RESULT 13  
US-09-807-007-6  
; Sequence 6, Application us/09807007  
; Patent No. 6881833  
; GENERAL INFORMATION:  
; APPLICANT: ZAPHIROPOULOS, Peter et al.  
; TITLE OF INVENTION: A NOVEL COMPONENT IN THE HEDGEHOG SIGNALLING PATHWAY  
; FILE REFERENCE: 2921-0130P  
; CURRENT APPLICATION NUMBER: US/09/807,007  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 6  
; LENGTH: 1447  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-807-007-6  
Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
Qy 225 PQGAVSGGIQPLNEGVARCNESQGDVATACSCODCAAPALARPALDSTFYLGMPSS 284  
Db 24 PERPAGGGRRRTGGIRRAAAPRD-----YLRPSYCDAAFALEGI--- 65  
Qy 285 LVLIILICSVPAVVTLLVGFVAPARDKSKMVDPK-----KGTSLSDKLSFSTHLL 337  
Db 66 -----SKGKATGRAPLMLRAKFORLLFKIGCYIQKNC 98

Qy 338 GQFPGMGWTAWSMPLTLLVSVIPVALLAGLVFTELTTPRVELMSAPNSQASEKAFH 397  
Db 99 GKF-----LVVGLLFGAFVAGLKAANLETIVBELWELVGVRSBELNYT 143  
Qy 398 DQHFEPFRNT--QVILTAPNRSRYRDBLLGPKNFSGILDLDELLE--LOERLRL 454  
Db 144 ROKIGSEAMFPQMLIQPKKEG-----ANVLTEALLQHLDBALQASRHHV 190  
Qy 455 QVMSPEAQRNISLODICYAP-----LNPNTSLYDCCINSLLQYFQNNRTLLLLTA 505  
Db 191 YMYN-----RQWLHEHCYKSGELITETGYMQDIIIEYLPCLILPDLDFMEGAKLQSGTA 246  
Qy 506 NOTLNGQ-----TSQYMDKHFLY-----CAN----- 527  
Db 247 --YLLGKPLRWMTNFDPLFLEELKKNYQVDSWEMLNKAEGVGYMDRCLNPAADDC 304  
Qy 528 -----APL-----TFKQGT-----ALALSCMA 544  
Db 305 PATAPNKNSTKPLDMALVNLGCGHGLSKRYHMQEELIVGTVKNSTGKLYSAHALQTMF 364  
Qy 545 DYGAIVFPFLAIGGYKDYSEAEALIMTFSLNNYPACDPRLAQAKWEAFLEEMRAFO 604  
Db 365 QLMTPKQWY---EHPKGYEY-----VSHIMNE---DKAAALIEAQRTYVEVHQSV 411  
Qy 605 RRMAGMFOVTTAERSLEDEINRTTAEPLPIATSYIYIFLYISLALGSYSWSRWVDS 664  
Db 412 AONSTQKALFTT--TTLDLILKSPSDSVIVASGYPILMLVACTM--LRMD--CSKS 465  
Qy 665 KATLGLGVAVVVLGAVMAAMGFPSTYIGRSLLVLIQVVPVLVSGADNIFLVEYQRL 724  
Db 466 QCAVLAIVLVALSVAAGLGLCSLIGISFMAATTOVLPFLALGVGVUDVILMAHAFET 525  
Qy 725 PRRPPEPRVHIGRALGVAPASMLCSISEAICFFLGLTMPAPVTFALTSGLAVIDF 784  
Db 526 GQNKRIIPREDRTGECLKTKTGASVALTISNTAFMALIPIPALRASFQAAVVVVENF 585  
Qy 785 LLOMSAPFALLSDBSKRDEARLDVCC-----VPRQEL-----PRG 822  
Db 586 AMVLLIFPAILSMDLXREDRLDIFCCFTSPCVSRVIOVBQVATTDHNTRRYSPPPY 645  
Qy 823 QCEGL----- 827  
Db 646 SSHAHEQTITMSTVOLRTEYDHTHYTTTAEPRSEISVQPTVTOTDLSQSPBST 705  
Qy 828 -----LIGFQKAVAPFLHWITRGVLLFLALFGVSL 861  
Db 706 SSTRDLDSQFSDSLHCLERPCWKMTLSSFAEKHAPFLKPKAKVVIFFLGLGVS 765  
Qy 862 YSMCHISVGLDDELALPKDSYLLDFLEFLNRYFEVGAVPYVTTLGYNSSSAGNAICS 921  
Db 766 YGTTVRDGLDLDIVPRETREVDFIAQFKYFSF-----YNN----- 803  
Qy 922 SAGCNNSFTOKIOYAT-----EPFQSYLAIPASS-----WVDFIDWL----- 961  
Db 804 -----YVTQKADYPNIOHLLYDLHRSFNKTYMLEENKOLPKMWLHYFRDMLOGLOD 857  
Qy 962 -----TPSS-----CRLYISGPNKDKFCSTVNSLNCNKMSITWGS 1000  
Db 858 AFDSMETGKIMPNNYKNGSDGVLAAYKLLVOTGSRDK--PIDISQLT--KQRLVDADGI 913  
Qy 1001 VAPSVQGFHKYL--PWFLND-----RPNIKCPKGGLAAY--STSVNLTSDGQ 1043  
Db 914 INPSA--FYIYLTAVNSNDPVAYAAASQANIRPHRPEWHDK--ADYMBETRLRIPAAR 968  
Qy 1044 VLASRFMAVHKPLKNSODYTEALRAARELANITA--DLRKVPOTDPAFEVPPYTTINVFY 1102  
Db 969 IEYAQFPFLNGLRDTSDEVEALEKVRTICSNYTSIGLSSYPNG-----YFP---LFW 1018  
Qy 1103 EGYLTLPBGLFMLSCLVPTFAVSCLLGLDLRSGLLNLISIMILVDVTFGMLMDIS 1162  
Db 1019 EGYGLRHWLLLFISVVLACTFLVCVAFPLNPMPTAGII--VMVLALMTVELFGMGLIGIK 1077  
Qy 1163 YNAVSLINVSAGMSVEFVSHITRSF--AISTKPTWLERAKEATISGSAVPAVAMTN 1220

Db 1078 ISAVPVIILASGIGVEFTVHVALAFLTAIDGN---RRAVLALHEHMFAPVLDG-AVST 1133  
Qy 1221 LGGILVLGAKQOLIOIFFFRLNLITLGLLHGLVFLPVILSYVG--PDVNPALALBQ- 1277  
Db 1134 LGLVLMAGSEDFDRIYFAVALITLIGVNLGLVLPVLSFFCPYFVSPANGLNRL 1193  
Qy 1278 --KRAEVAVAWVASCNNPSRVSTADNIYVNHSPESIGK 1317  
Db 1194 PTPSPPPSVVRFPMPFGHTH--SGSDSDSEYSQTTVSG 1233

RESULT 14  
US-09-754-032-19  
Sequence 19, Application US/09754032  
Patent No. 6921646  
GENERAL INFORMATION:  
APPLICANT: SCOTT, MATHEW P  
GOODRICH, LISA V  
JOHNSON, RONALD L  
TITLE OF INVENTION: Patched Genes and their Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/754,032  
FILING DATE: 03-Jan-2001  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/540,406  
FILING DATE: 06-Oct-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertram I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: a60190-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1447 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-754-032-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

Qy 225 PGQAVSGIOPLNMGVARGNESQGDVATCSQDCASCPAIAAPQALDSTFYLGQWPGS 284  
Db 24 PGRPAGGGRRRRGGLRRAAPDRD-----YIHRPSYCDAAFALEQI--- 65  
Qy 285 LVLIILICSVFAVVTLLVGRVAPARDKSMVDPK-----KGTSLSDKLSFSTHTLL 337  
Db 66 -----SKKATGCRKAPLMLRAKFGQLFLKGLGCIQKNC 98  
Qy 338 GQFFQMGKTVWASWMLTILVLSVIVVLAAGLVFTLTDPEVLMSAPNSQARSEAFH 397  
Db 99 GKF-----LVVGLLIFGAFVGLKAAMLETNVELMVEVGGRVRSREINMT 143

Qy 398 DQHFQFFRRTN-QVITLTPANSSRYRDLGLGPKPSSGILDDLLLELLE--LQERLRL 454  
Db 144 RQKIEEAMFNQOLMIQPKKEG-----ANVLTEALLQHLSDALQASRVH 190  
Qy 455 QVWSEADQANISLDQICAP-----LINDNTSLVDCCINSLSLOFQONRTLLLLTA 505  
Db 191 YWYN-----RQWKLHLCYKSGELITETGYMDIILEYLPCLLITPLDCTWBAKIQSGTA 246  
Qy 506 NQTLNGQ-----TSQVMDKHFLY-----CAN----- 527  
Db 247 --YLIGKPPRLRWTPDPLFLEELKKINQVDSMEEMKAEVGGYNDRPLCLNPRDPDC 304  
Qy 528 -----APL-----TFKQGT-----ALALSCNA 544  
Db 305 PATAFNKSTKPLDMLALVINGCHGLSKRMHMOBELIVGGIVNKSTGLVSAHMLQTFM 364  
Qy 545 DYGAIVFPFLAIGYKGDYSEAEALIMTFSLNNYPADGPRLAQKLEAEFLKEKRAFO 604  
Db 365 QLMTPRKQWY---EHFKGYEY-----VSHINNE---DKAAAILAEMQRTYVEVHQSV 411  
Qy 605 RRMAGMFOVTFPAERSLEDEINRTTAEDLPFATSYIVIFLYISLALSGYSSMSRVNDS 664  
Db 412 AQNSTQKVLSTFTT--TLLDDILKSFSDVSIVRVASGILMLAVALCLTM--LRMD--CSKS 465  
Qy 665 KATLGLGVAVVGLGAMAMGFPSYLGIRSSLVILQVPELTVSGADNIFIFVLEYQRL 724  
Db 466 QGAVGLAGVLLVALSVAGLGLCSLIGSENAATTQVPELALGVGDVDFLAAHAFSET 525  
Qy 725 PRRPGPREVHIGRALGRVAPSMGLCSLSEALCFIFGALTTPMAVETPALTSGLAVLDF 784  
Db 526 GQMKRIPEEDRTGECIKRTGASVALTSISNVAFFMAALIPFALRAPSLOAAVVVVFNF 585  
Qy 785 LLOMSAFVALLSLDSKROQASRLDVCCC-----VKQDEL-----PPG 822  
Db 586 ANVLIFPALISMDLYRRDRDLDFCCTSPCVSRVIGVEQATYDTHDNTRYSPPEY 645  
Qy 823 QGEGGL----- 827  
Db 646 SSHSFAHETQIMQSTVOLRTEYDPHTHVYTTAEPRESEISVQPVTVTQDTLSCQPEST 705  
Qy 828 -----LGFPOKAYAPFLIMHTRGVULLFLALFGVUL 861  
Db 706 SSTRDLLSQFSDSLHCLPCTKWTLSGFAKHVAPFLIKRAKVVVIFLGLGIVSL 765  
Qy 862 YSKCHISVGLDQELALPKDSYLLDYFLNKRFEVCAVYFYTYLGYNSESSEGMALIS 921  
Db 766 YGTTVRVDGLDLTDIVPRETREVDFIAQKTFYSF-----YNN----- 803  
Qy 922 SAGCNNPFTQKIQVAT-----EPPEGSYLAIPASS-----WVDFIDPL----- 961  
Db 804 -----YIVTQKADYVNIQHLVLDHRSFSNVKTYVMLEENKOLPKMMLHYFRDMLQGLD 857  
Qy 962 -----FPSS-----CCRLYISGPNKDKFCPTVNSLCLKNCMSITWGS 1000  
Db 858 APDSMETGKIMPNYKNKSGDDGVLAAYKLLVOTGSDK--PIDISGLT--KQRLVADAGI 913  
Qy 1001 VRPSVEQFHKYL-PWFLND-----RPNIKCPKGLAAY--STSVNLTSDQO 1043  
Db 914 INPSA--FYIYLLAWNSNDPVAAAGANIRPRHPWVNDK--ADYMETRLRIPAAEP 968  
Qy 1044 VLASRFMAHYKPKANSQDYTEALRAARELANITA-DLAKVPQTDAAFEVFPPTINVPY 1102  
Db 969 IEVAQFPFYNLGJRDSDEVEALEKRTICSNYTSIGLSYSPNG-----YFP-----LFW 1018  
Qy 1103 EQVLTILPBGIFMLSLCIVPTFAVSCLLGLDRLSGLMNLISVMTLVDPVGMALMDS 1162  
Db 1019 EOYIGLRHMLLLEFISVLAICTFLVCAVFLINPTAGIT--VMVALMTVELFGMGLIGLK 1077  
Qy 1163 YNAVSLINIVASVAGVEFVSHITRSF--AISTKPTMLBRRAKATISMSGAVPAGVAMTN 1220  
Db 1078 LSANVPVILIASGIGVEFTVHVALAFLTAIDGN---RRAVLALHEHMFAPVLDG-AVST 1133





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using bw model

Run on: March 22, 2006, 22:51:02 ; Search time 53 Seconds  
(without alignments)  
2418.127 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909  
Sequence: 1 MAEAGLRGWLWALLRLRLA.....GSIKGAISNPLPNNRGQF 1332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2385	34.5	1278	2 T30188	Niemann-Pick C dis
2	1394	20.2	1055	2 T05663	hypothetical prote
3	1347	19.5	1456	2 T15961	hypothetical prote
4	1329	19.2	1170	2 S52525	probable membrane
5	646.5	9.4	1434	2 T30172	transmembrane prot
6	628.5	9.1	1442	2 T18538	patched protein -
7	600	8.7	915	2 S44797	F0968.4 protein -
8	598.5	8.7	933	2 T25600	hypothetical prote
9	583.5	8.4	1182	2 T13952	membrane protein p
10	573.5	8.3	1003	2 T26746	hypothetical prote
11	572	8.3	1220	2 T18291	patched protein -
12	539.5	7.8	889	2 T29590	hypothetical prote
13	514	7.4	1405	2 T27969	hypothetical prote
14	473.5	6.9	1015	2 T15830	hypothetical prote
15	464.5	6.7	1299	1 S06119	membrane protein p
16	458.5	6.6	955	2 T21612	hypothetical prote
17	436	6.3	956	2 A89153	protein C24B5.3 [1
18	409.5	5.9	936	2 T26521	hypothetical prote
19	407	5.9	800	2 T26683	hypothetical prote
20	399.5	5.8	881	2 T31739	hypothetical prote
21	381	5.5	840	2 T33217	hypothetical prote
22	379	5.5	714	2 T23399	hypothetical prote
23	367	5.3	690	2 T23399	hypothetical prote
24	367	5.3	877	2 T24097	hypothetical prote
25	349	5.1	845	2 T25657	hypothetical prote
26	347.5	5.0	820	2 T32908	hypothetical prote
27	312	4.5	890	2 T22186	hypothetical prote
28	297.5	4.3	871	2 T28706	hypothetical prote
29	294	4.3	983	2 T21213	hypothetical prote

30	276.5	4.0	633	2 S44795	F0968.3 protein -
31	259.5	3.8	413	2 S28276	hypothetical prote
32	224	3.2	1276	2 T18526	SREBP cleavage act
33	189.5	2.7	1154	2 T48829	related to SREBP c
34	187.5	2.7	932	2 T28820	hypothetical prote
35	186.5	2.7	1227	2 T20370	hypothetical prote
36	179.5	2.6	787	2 H71453	hypothetical prote
37	173	2.5	969	2 T33156	hypothetical prote
38	170	2.5	746	2 A75018	transport protein
39	154	2.2	823	2 B81282	probable integral
40	149	2.2	1011	2 T07712	probable ABC-type
41	149	2.2	1051	2 A93455	multidrug resistanc
42	148.5	2.1	746	2 H84301	hypothetical prote
43	146	2.1	1041	2 AC0423	multidrug efflux p
44	143	2.1	724	2 H69780	antibiotic transpo
45	141.5	2.0	1049	2 AF0561	acriflavin resistanc

ALIGNMENTS

RESULT 1

T30188  
Niemann-Pick C disease protein - mouse  
N:Alternate names: NPC1 protein  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T30188  
R:Loftus, S.K.; Morris, J.A.; Carstea, E.D.; Gu, J.Z.; Cumming, C.; Brown, A.; Ellison, Science 277, 232-235, 1997  
A:Title: Murine model of niemann-pick C disease: mutation in a cholesterol homeostasis A:Reference number: 220765; PMID:97362324; PMID:9211850  
A:Accession: T30188  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1278 <LOF>  
A:Cross-references: UNIPROT:Q35604; UNIPARC:UP100000295BC; EMBL:AF003348; NID:G2251241; C:Gene: NPC1  
A:Gene: NPC1  
A:Map position: 18

Query Match 34.5%; Score 2385; DB 2; Length 1278;  
Best Local Similarity 40.2%; Pred. No. 1.2e-158;  
Matches 525; Conservative 233; Mismatches 443; Indels 104; Gaps 27;

QY	14	LLRLAQSPPYTHIQPGICAFYDECCKNPBELSGSLMTLSNVSCLSNTPARKITGDHLTL	73
DB	12	LLRLCPAQVFSQ-----SCVMYGCQ--LAVGD--KRYNCKYSGPKPLPKDGYDL	59
QY	74	LQKICPRLVYTGNTQACSAKQVLSLEASLSTTKALTRCPACSDNFVNLHGNTCSPNQ	133
DB	60	VELCPGLFF-DNVLSCDICIQDLQTKSNLQPLQPLSLCPCPCFNLMTLPELITCSPHQ	118
QY	134	SLFINVTRVAQAGQLPA---VVAEAFYQHSFAEODSYDCSRRVRPAALTLAVGTC	189
DB	119	SEFLNVTATEDYEDFPKTPENKTNKELEYVGCSPANAMYNACRDVEAPSSNEKALLGC	178
QY	190	GYVGSALCNAQRMWLNFGSTGNGLAPLDI-----TFHLLPEQNAVSGIQPLNEGARCN	244
DB	179	GDARA-CNATWIEFMKDKNGQAPFTIIPVPSDLSTL-----GMEPRNATKGN	229
QY	245	ESQGDVATSCGDCCAASC-----PAIRPOLADSTF-----YGGMGSLV	286
DB	230	ESYDEVYTGSCODCSYVCGPRPQPPPPMPRIRINGLDMMYIMVTVYAFVFPFGALL	289
QY	287	LIIILCSFAVVTILLVGRVAPARDKSNVDPKKGTSLSDYLSSTHTLLQFQSGCT	346
DB	290	AWCHRRRYFVSEYTPIDISNIAFSVNS---DKGEASCCDPLGAFFDDCLRMFTKWA	345
QY	347	WASNPFLTLVLSVFPVVALAGLVFTTELTTPVPLWMSPNQASSEKAFHOHGPFR	406
DB	346	FCVRNPCTIIFPSLAFTVCGSLVFPVQVTVNPVELMSAPHQARLEKEYPKHGPFPR	405

QY 407 TNOVLTANRRSSRYDLSLLGPK-NFSGILLDILLLELEOERLRHQVSPENQR- 464  
DB 406 TEOLIIQAVENTSVHIEPEYPAGADVFPGLNKEIHHQVLIHQ-----IAISITASYNN 460  
QY 465 --ISLQDICYPALNPNTSLVYOCCTISLQYFONNRTELLLTANQTLMGOTSOV----- 516  
DB 461 EYVTLQDLCVABLSPPYKN--NCTIWSVLTNYFONSNAVL-----DSQVDDPFI 506  
QY 517 --DWKDHPLVCANAPLTFKDGTAALASCMADYGAPVPEPLAIGYKGDYSEAEALIMTF 574  
DB 507 YADYHTHFLYCVRAPASLNDTSLHGPCLGTGCGPVPMLVLYGVDQDQYNNATALVITF 566  
QY 575 SLANTPAGPRLAQAQCLMEAEFLBEEKRAFORRMAGFOVTFTRERSLEDEINRTTADLP 634  
DB 567 PNNVNTYDTERLQRAWAMEKEFISFYKNYKN--PMLTISFPAERSIEDELRESNDVF 623  
QY 635 IPATSYIVIFLYISLALGYSYSSWRVWDSKATLGLGVAVVLGAVMAMGFFSYLGIRS 694  
DB 624 TVIISYVMEFLYISLALGHIOGSCRLLVDSKISLGIAGILYLSVACSLGIFSTYGMFL 683  
QY 695 SLVILQVVPFLVLSVGNINIFIVLEIYQRLPRRPGSPREVNIGRALGRVAPSMILCSLSE 754  
DB 684 TLIVIEVIFPLVAVGVNDNIFILQYQORDERLQEBTLDQGRILGEVAPTFMILSSFSFE 743  
QY 755 AICFPLGALTMPAYARTFALTSGLAVILDFLLQMSAFVALLSDSKRQASRLDCCYK 814  
DB 744 TSAFFFGALSMAPAVHTFSLPAGMAVLIDFLQITCFVSLIGLIDIKRQKNDILICVR 803  
QY 815 PQLPPLPGOG----EGLLGPFQOKAYAPFLHMITGVLLLFILAFGVLSXSMCHISVG 870  
DB 804 GAD--DGGSHASESYLFRFFKNYAPILDKMLRPYVAVFVGLSSVAVNKKVDIG 860  
QY 871 LDQELALPKDSYLLDYFLFLNRYFEVGAPVYFVTLTGYNFSEBAGNNAICSSAGCINPFSF 930  
DB 861 LDQSLMPRDSYVIANFKSLAQYLHSGRPVYFVLEEGYNYSSRKQNMVCGMGCONDBL 920  
QY 931 TQKIQYATFEPEOSYIATPASSWVDFITWLP--SSCCRLYISGPKKDFCSTYNSLNC 989  
DB 921 VOQIFMAELDYTRTGFAPSSWIDYFDMVSPQSCRLY--NVTHQFCNASVMDPTC 977  
QY 990 LKNCMSIT-MGSVRPVEQFHKYLPWFLNDRPNIKPKGGLAAYSTSVNLTSQGYLT-AS 1047  
DB 978 VR-CRPLTPEGKORPGKGFPMKFLPFLSDNPNPKGKGGAAYGSAVIVGDDTYIGMT 1036  
QY 1048 REMAYHKPLKNSQDYTEALRAARELANITADLRKVPCTDPAFEVPPYITTVFEYQYLT 1107  
DB 1037 YEMTYHTILKTSADYTDAMKAKRLIASNITEMWRS-KGSD--YRVPYSVFVFEYQYLT 1093  
QY 1108 ILPEGLFMLSCLVFPFANVSCLLGLDLSGLNLSTYMLVLDYTVGFPAALDISYNNAS 1167  
DB 1094 IIDDITFNLVSLSGISFLVTLVVLGCELMSAVIMCTITAMILVNMFGVMLGSLNAVS 1153  
QY 1166 LNLVSAVMSVEFVSHITRSPFAISTKPTMLERAKATISMGSAVPAVAMTNLPGLIYL 1227  
DB 1154 LVNLVMSGISIVBFCHITRAFLTMSYKSGRVSRAEALAHMSSSVSGITLTKFGSIVVL 1213  
QY 1228 GLAKAQLIQIFFRILNLITLLGLLGLVFLPYILSYVDPVNP 1272  
DB 1214 ARAKSGIFEIFYFRMYLANVLLGATHTGLFLPLVLSYIGPSVKNKA 1258

RESULT 2  
T05663  
hypothetical protein F22113.120 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05663  
R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15420  
A:Accession: T05663  
A:Molecule type: DNA  
A:Residues: 1-1055 <BEV>

A:Cross-references: UNIPROT:O95VF0, UNIPARC:UP100000A552EC, EMBL:AL035539  
A:Experimental source: cultivar Columbia; BAC clone F22113  
C:Genetics:  
A:Map position: 4  
A:Introns: 24/3; 60/3; 99/3; 150/1; 193/1; 216/3; 278/3; 297/2; 336/3; 364/3; 396/3; 41  
026/3  
A>Note: F22113.120

Query Match 20.2%; Score 1394; DB 2; Length 1055;  
Best Local Similarity 28.5%; Pred. No. 2,6e-89;  
Matches 377; Conservative 221; Mismatches 385; Indels 340; Gaps 41;

QY 36 YDECGKNEBLSMTLSNVGCLSTNPARKITGDHLI--LLOKICPRLYGTENQACCSA 93  
DB 2 YDICHRSDD--GKVL-----NCPYASPS--IQPDELFAKIQSLCPTI-----SGNVCTE 48  
QY 94 KQLVLEASLSTIKLTLTRCPACSDNPNLHCHNTCSNOSLFINVTRVAQAGOLPAV 153  
DB 49 TQFDLTRSQVOQAAPFLVGCRCALRNFLNLFCELSCSFNOSLFINVTVAAV--SGNL-TV 106  
QY 154 VAYEAFYQHSFAEQSYDSCSRVRVPAATLAVGTWGVYSALCNAQRMFLPQD--TG 210  
DB 107 DGIDYHITDTFGEGLYESCKEYKFGTMMTRAINFV---GGAKNFRMFTFICQKAPSG 162  
QY 211 NGLAPLDTFHLBFGQAVGSGIQPLNCGVARNCHESQGDVATGSCQCCAAACRAIARQ 270  
DB 163 PFGSPYAINFSSSLP---ESSAMPMN-----VSYSACSSPEPLPPI 203  
QY 271 ALDS--TFYLG-----QMPGSLVLIILIGSAFVAVTLLVGFRAPARDSKNVDPKKG 322  
DB 204 DEBSCSITIGPLKRCIELSMALYVLLVSCFEGMAGNRRNRTOPIDSKRPLHPBE 263  
QY 323 TSLSDKLSFSTHTLLG-----QFQCGMTVNASWPLTILVSVIPVA 365  
DB 264 DQINSEMK---ENILGVKVRHAQLSPVQRYMAKYRSYSGMIARNPFLVFMVAIVLA 320  
QY 366 LAAGVFLPBLTDPPELMSAPNSQARSKAHDOFGFPFTNQYIL-TANRRSYRDS 424  
DB 321 LCGSYLNKRVETRPBKLVWGESKAEKKEFDTHLSPFYRIEOLILATVDDPKSGRPS 380  
QY 425 LLLGPKNFSGLIDDLLELEOERLRHLQWSPQASRLSLQDICYAPLNPDTSLYD 484  
DB 381 I-----VTDEHILL-LPDIQK----- 396  
QY 485 CCINSLQYFONNRTELLLTANQTLMGOTSOVDMKHEFLVCANAPLTFKDGTAALSCMA 544  
DB 397 -----YFKMD-----SGTFDDYGVGEHAEC-----FOHYTS-SETCLS 429  
QY 545 DYGAVPFPFLAIGYKGDYSE-----AALIMTFSLANYP 580  
DB 430 AFOAEVVDPSAVLGGFSGNYSBWWVSELGCVPEDCYSDVKRTLFQATAFVVTYVNVN- 488  
QY 581 AGDP--RLAQACLMEAEFL-----EEMRAFORRMAGFOVTFTRERSLEDEINRTTADLP 634  
DB 489 IGDSSNEAKRAVAMAKSRIQLAKEELLPMVR--SKNLSLSTSSBSIESBELKREBTAVI 546  
QY 635 IPATSYIVIFLYISLALGYSYSSWRVWDSKATLGLGVAVVLGAVMAMGFFSYLGIRS 694  
DB 547 TIAASYLVMEFYISVTLDDAPQFYFYISSKVLGLSLGVVLVLSLVGVSQVFSALGVKS 606  
QY 695 SLVILQVVPFLVLSVGNINIFIVLEIYQRLPRRPGSPREVNIGRALGRVAPSMILCSLSE 754  
DB 607 TLIVIEVIFPLVAVGVNDNIFILQYQORDERLQEBTLDQGRILGEVAPTFMILSSFSFE 743  
QY 755 AICFPLGALTMPAYARTFALTSGLAVILDFLLQMSAFVALLSDSKRQASRLDCCYK 814  
DB 665 VLAEPVGAFPVMPACKRITSMFRAALAIMDFLQITAFVALLVFDCKRSADRIRIDCFPIK 724  
QY 815 ----PQLPPLPGOGEGLLIGFQOKAYAPFLHMITRGVALLFLALFGVLSXSMCHISVG 870  
DB 725 VPSSRSRESVEGRREGFLERYMKVHAIVLGLMGVKNVAVVAFV----- 769  
QY 871 LDQELALPKDSYLLDYFLFLNRYFEVGAPVYFVTLTGYNFSEBAGNNAICSSAGCINPFSF 930

Db 770 -----FAL 772  
OY 931 TOKIYATEPPEOSYLAIIPASSWVDDFIDMLTPSS--CCRLYISGNKDKFCSTVNSLN 988  
Db 773 ASISPAASQASDSTSIYAKPAPASWDDFLVWLSPEAGCCCKFTNG-----SYCP--DDQ 825  
OY 989 CLKNCMSITMGSVRPSVEQFHKYLPWFLNDRPNKCPKGGLAAYSTSVNLT--SDGVOLA 1046  
Db 826 CFFHS---DLVQGRPSTAGREKLPWFNLNLPADCAKGGHGAVTNSVDLKGVESGVIOA 882  
OY 1047 SRPMVHKPLKXSQDYTEALRARELANITADLRKVPGRDPAEVFPPTITNVFEQYL 1106  
Db 883 SEERTHTPL-NTQ-----IDIFPSVVFYIFFEQYL 912  
OY 1107 TLPEGLFMLSCLVLPFAVSCLLGLDLSGSLNLSIWMILVDTVGFMAIDISYNNV 1166  
Db 913 NITWVNLTLNLAIV-----GIQLNAV 934  
OY 1167 SLINLVSAGMSVEFVSHITRSPASTKPTMLERAKATISMGSAVAGVAMTNLPGLIV 1226  
Db 935 SYVNLMSIGIANEFCHISHATLMSGDR-BHRARBALETWGAASFSGITLTKLVGYIV 993  
OY 1227 LGLAKQQLQIFFFRNLTLTLGLHGLVFLFYIISYVG--PDVNPALAEQKAEANA 1285  
Db 994 LCFARSEIFVYVYFQWYALVILIGFLHGLVFLFYIISLAGPPQIN--LDIEQOOTDEASS 1051  
OY 1286 AVN 1288  
Db 1052 SLL 1054

## RESULT 3

T15961

Hypothetical protein F02B8.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_rev:1510 20-Sep-1999 #ext\_change 09-Jul-2004

C:Accession: T15961

R:Miller, N.

Submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F02B8.

A:Reference number: Z18436

A:Accession: T15961

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Gene: CRSP.F02B8.6

A:Experimental source: strain Bristol N2; clone F02B8

A:Map position: X

A:Introns: 71/1; 124/3; 218/3; 269/3; 310/3; 403/2; 484/3; 528/2; 764/1; 858/3; 888/3; 1

Query Match 19.5%; Score 1347; DB 2; Length 1456;

Best Local Similarity 27.5%; Pred. No. 7.7e-86;

Matches 372; Conservative 258; Mismatches 565; Indels 158; Gaps 41;

OY 76 KICPRLYTGPNYACCSAQVLSLESLSTTKALLRCPACSNFPNHLGHNTCSNQS 135  
Db 63 EFCPHLLTGDN-KLCTPSQAEGLTKIOAQRHILCRCPSCFDNFAKLWCEFTCSNPOD 121  
OY 136 FINTVVAQL--GAGQLPAVVAEAF-----YQHS--FAGQSYDSCSRVAPAAATLAVG 186  
Db 122 FVSIEMKPIKEKGETPEYQPAEAYVNYEYLSLDFABGMSCKDDTVFGQPALRV- 180  
OY 187 TWCYVGSALCNARWMLNFGDGTNGU-APLDITFHLEBPGAVGSGIOP-LNEGVARCN 244  
Db 181 -WC---TSFPCITLTMWLEFIGNLNLNIPHTKFLLYDPIKTPPSDRSTYVMNVFTGCD 236  
OY 245 ESGDDVATCSGDDCAASCAPARFQALDSTFLGCM-----PGSLVLIILLCISVAVT 239  
Db 237 KSARVGMWPACTSEC--NKEEYANLIDLDDGKTSGGTGNVHGIACTINFVMAFIQSLAV 294

OY 300 ILLVGF-----RVAPARDXSKWVDPKKGTSLSDKLSFSTHTLLGQSF 341  
Db 295 LICEVFVFTSYDEDTYNLRQTOSESSPKGNRIK--RTGAMI-----HNFEMENNA 342  
OY 342 QCMGTWVAMWPLTLLVLSVIVVVALAAGLVFELTTPDELMASAPNSQARSEKAFHDQF 401  
Db 343 RDIQMAGNPNPSSHFTIGCAVILFCLPGMTYHKESTNVVDMSSPSSRLROEEMVNNAP 402  
OY 402 GPEFRTQVILTAPNRSRYRDLNLLGPNFSGILDLLELLELQERLRLQVNSPEA 461  
Db 403 GRPQRQOIML--SHRDFQSSGKLXG-----VFHKDIFELFDILNMIKNISTQDSQ 455  
OY 462 QRNISLQDICYAPLNDNTSLDCCNSLLQTFQNNRTLLLTANQ-----L 509  
Db 456 -RTITLDVCCYRPMGCG---YDCLMSPTNFOGKEHLDKSNKEETVSEDDDAFDYF 510  
OY 510 MGQTSQVMDKDFLVCANAPLFPKQDATALSCMDYGAVPFPELAGYKQDYSEAEV 569  
Db 511 SSBATIDEMNHMAALIDQPMQOK--TKSGLSCMGTYGSPAPNM--VFQKNSITNHQAANS 567  
OY 570 LMTFSLANNYPAGDPELAQAKLMEBAFLEMBRQFRMAGMFQVTFARSLDEINRTT 629  
Db 568 IMMTILVQ--RTBPEIQAKELMEKEFLKCEYREKSPKVI-FSPMAERSITDEIENDA 624  
OY 630 AEDLPFATSYVIFLYISLALGSY-----SSSRVMDSKATLGLGVAVVLGAVMAAM 684  
Db 625 KDEIVTVVIALAFLIGVYFSLGRYFVCENQLMS--ILVHSRICTGLMSYIINLSSFCGW 683  
OY 685 GPFSTYIGRSSVLIVQVFPVLVSGADNIFIFVLEY--QRLP---RRGEPREVIIGSA 739  
Db 684 GIFSMFGHIPVKNALVVOFFVVTLLGVCRFVWVKRYAQQRVMPYMSDQCEI-VGAV 742  
OY 740 LGRVAPSMLLGSLSEALCFELGALPMPAPRTFALTSGAVLILDLQSAFAVALSLDS 799  
Db 743 MAGTMAPMSSSLGCAFSPFIFGFTDLPRIKFTCLYAGLAVLIDVLAHCTITPLALFVMT 802  
OY 800 KQGEASRLDCCVCKPQELPPQGGEL--LLG-----FQKAVAPFLHMI 844  
Db 803 QRELNG-----KR-EFFFPYQIKDLGAYLIGORATDFEMTQFFHFQVAPFLMHM 853  
OY 845 TKGVULLEFLALFGVSLYMKCHISVGLDDELALPKDSYLLDFFLNRFVBCAPRYFT 904  
Db 854 TRIIIGIPIAFITTVIILSSKISVGFDDSMKFTKSYSTHPRYDKDFDGPVFPV 913  
OY 905 TLGVNFSBAGNNAIISAGCNNSFTQKIQYATEPPEOSYLAIIPASSWVDDFIDMLT-P 963  
Db 914 DGEIDHNRDVOVKKFTFFGCGSDTSGNIMNTAVAGHTEQTYLSGEMYMINDYLEMISRK 973  
OY 964 SSGCRLYISGPNKDKFCPSTVNSL-----NCLKNCM-----SITWGSV---RPSYEQFH 1009  
Db 974 SPCCKYYVADPN--FTCSITNRKMSALDKACRTCMDFDVANSYPKSSIMYHRPSIEVFY 1031  
OY 1010 KYLPWFLNDRPNKCPKGGLAAYSTSVNLTSQCVLASFPMAVHKRL--KNSQDYTEAR 1067  
Db 1032 RHLRHELEPTNSECEVGRASFKOAIISFISGRIOASQFMFHKKLSISNSDFIKAND 1091  
OY 1068 AARELANITADLRKVPGFDPAEVFPPTITNVFEYQYTLIPBGLFMLSCLVTPFAVS 1127  
Db 1092 TABMWSRRLERSI-----DRTAHVFAYSKIPFIEQYSTIMPILTQQLFITYVSGFGLI 1145  
OY 1128 CLLGLDLSGSLNLSIWMILVDTVGFMAIDISYNAVSLNLSAVAGMSVEFVSHIR 1187  
Db 1146 CVTLGIDVGAACAVICQVSNFHLVAFMYIFINIPNALSATNVLVMSGILFEFSVNLK 1205  
OY 1188 SFAISTKPTWLBRAKEATISMGSAVAGVAMTNLPGLIVLGLAKQLQIIFPRRLNLLT 1247  
Db 1206 GYACSLRQRAKDRABESTVSGIPILISGVVIMAGSTWMLSGAHQIITVYFQKFLIYI 1265  
OY 1248 LGLHGLVFLFYIISYVGPVDNPPALALEQKAEAVVAVMVASCNPHSRSTADNIIY 1307  
Db 1266 VSSAVHALIILPILALFGGSRGSGSETSTNDNDQHDACVLS--PTASHISNVBEGIL 1323  
OY 1308 N-----HSFEGSIKGAISNPLPNNGRQF 1332

Db 1324 NRPSLIDASHILIDPLIKAEGLDKAI--GRDF 1353

RESULT 4  
S52525  
probable membrane protein YPL006w - yeast (Saccharomyces cerevisiae)  
N/Alternate names: hypothetical protein Lpaliw; hypothetical protein YP8132.07  
C/Species: Saccharomyces cerevisiae  
C/Date: 08-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C/Accession: S52525; S59687  
R/Badcock, K.; Churcher, C.  
submitted to the EMBL Data Library, February 1995  
A/Reference number: S52519  
A/Accession: S52525  
A/Molecule type: DNA  
A/Molecule type: DNA  
A/Residues: 1-1170 <BAD>  
A/Cross-references: UNIPROT:Q12200; UNIPARC:UPI000006A19C; EMBL:Z48483; NID:G683777; PID  
A/Experimental source: strain AB972  
R/Hall, J.; Ahmed, A.; Busey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; We  
submitted to the EMBL Data Library, August 1995  
A/Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.  
A/Reference number: S59677  
A/Accession: S59687  
A/Molecule type: DNA  
A/Residues: 1-1170 <HAL>  
A/Cross-references: UNIPARC:UPI000006A19C; EMBL:U33335; NID:G965076; PID:G965087; MIPS:Y  
C/Genetics:  
A/Gene: SGD:NCR1  
A/Cross-references: SGD:S0005927; MIPS:YPL006w  
A/Map position: 16L  
A/Keywords: transmembrane protein  
C/Keywords: transmembrane #status predicted <TM1>  
F.1-1/Domain: transmembrane #status predicted <TM2>  
F.258-274/Domain: transmembrane #status predicted <TM3>  
F.341-357/Domain: transmembrane #status predicted <TM4>  
F.585-601/Domain: transmembrane #status predicted <TM5>  
F.614-630/Domain: transmembrane #status predicted <TM6>  
F.667-683/Domain: transmembrane #status predicted <TM7>  
F.698-714/Domain: transmembrane #status predicted <TM8>  
F.751-767/Domain: transmembrane #status predicted <TM9>  
F.1004-1020/Domain: transmembrane #status predicted <TM10>  
F.1027-1043/Domain: transmembrane #status predicted <TM11>  
F.1051-1067/Domain: transmembrane #status predicted <TM12>  
F.1103-1119/Domain: transmembrane #status predicted <TM13>  
F.1137-1153/Domain: transmembrane #status predicted <TM13>

Query Match 19.2%; Score 1329; DB 2; Length 1170;  
Best Local Similarity 28.8%; Pred. No. 1.1e-84;  
Matches 373; Conservative 230; Mismatches 512; Indels 182; Gaps 46;

QY 10 LLMAL-----LRLAQSEPYTHHPGCAFYDECGKNP-----ELSGSLMTLSNVSLSN 60  
Db 3 VLMITLVGQIMRLVGG-----TATCAMYGNCKSKSVFQNELPCPRSPRPV 54  
QY 61 TPARKITGHLILQKICRPLTYGPNTOACCSAKQVLSBASISTKALLTRCPACSDNF 120  
Db 55 ETSK-----LTVCEBEMKEVR-YACCTKDQVALRDMLOKQAPLISSCPALCKNF 105  
QY 121 VNLHCHNTGSPNQSILINTTRVAQOLGAGQPAVVAEAFQHSFAROSYDSCSRVAVPA 180  
Db 106 NNLFCHFTCAADGRVNVITKV-EKSKEDKDIYAEIDVFANSSMAEFIDSCNINFSAT 164  
QY 181 ATLAVGTMGVYGSALCNARMLNFOGDTGN--GLAPLDITFHLRPGQAVSGGIQPLNE 238  
Db 165 N-----GYANDLIGCGAKNYSQFLKFLGDAKPMLGSGSPFOINXYKDANBE--KEWGEFND 218  
QY 239 GVARCHESGQDVATSCQDCAASCPAIAAPQALDSTFVLGQMPG---SLVLIILICSVF 295  
Db 219 EYVACDDAQ---YKCAQSCQSCPSPHL-KP-LKDVCKVGPJLPCPSLVLIIFYTICALF 272  
QY 296 AVYTTILVGFVAPAPADKSKMVDPKKGTSLSDK-----LSFSHT-----LIGQFQGWG 345  
Db 273 AFWMYTLCKRKKGAKMIYDDDIVPESG-SLDESETNVFESFNNEITFENGKLANLFTTKVG 331

QY 346 TWVASMPLTILVLSIVPVALLAGLV-FTELTPDVEIWSAPNSQARSEKAFHDOHFGPF 404  
Db 332 QPSVENPYKILITTVFSIFVFSFIIFOYATLETDPINLWNSKSEKFEKEYEDDNFGPF 391  
QY 405 PRTNOVILTPARRSVRYDSL--LLGPNFSGILDLDLLELLEQLERHLHQVNSPEQ 462  
Db 392 YATGEQIFVWNETGPVLSYETLHMFDVENF-----ITBEL-----SS 429  
QY 463 RNISLQDICYAPLNDPNTSLYDCINSLSLOYFQNNRTLLLTANQTLNGQTSQVMDKH 522  
Db 430 ENIGYQDLCFR-TDST-----CVIESFTQYFQG-----ALPKDSMKEL 470  
QY 523 LYCANAPLTPKDGITALASCADYGAVPFPLAIGYIKDYSEAEALIMPSLNNYPAG 582  
Db 471 QECGRFP-----VNCLEPTFOQPLKTNLL--FDDDDLNNHAFVVLTLTNH-- 514  
QY 583 DRLAQALMEBAFLSEMRAPGRMAAGFOYFTFERSLEDEINNTAEDPIFATSYIV 642  
Db 515 ---TOSANRMEER-LEBYLLDKVEBGL-RISFNTLSLEKELANN--NDISTVALSYLM 567  
QY 643 IFLYISLALGSYSWSRWVDSKATLGLGVAVVLGAVMAAMGPFSSYLGRSSVILQV 702  
Db 568 MFLVATMALARDKDGRLL-----LGISGLIVASIVCAAGFLTLGLKSTLITAEVI 621  
QY 703 PFLVLSGADNIFIFVLEIQR-LPRRQEPREVAHIGRLGVRAPSMLLCSLEAICFLG 761  
Db 622 PFLIILAIGDNIFFLTHERYDRNCEQKPYPSIDOKLISAIGRSPSILMSLQCQTCFLIA 681  
QY 763 ALTPPAPVTRPALTSGLAVILDFLOMSAFVALSLSDKQDASLDVCCVQPELPP 821  
Db 682 AFVTMPAVHNFATYTSVIFNGVQLTAYVSIISLYEKSRYKOT----- 728  
QY 822 GQGEGLLGFQFKAPAPFLHMTGRVVLFLFALFGVLSYMSCHISVGLDQELAPXDS 881  
Db 729 -GNEETKSEFLKTFYFKMLTQ---KRLIIIFSAFFSVLPLPIQIGDQTLAVPDQS 784  
QY 882 YLIDYFLFLNRYEYGAQVYFYTTIGVNPSSBAGNNAICSS-AGCANNFSFTQKQYATEF 940  
Db 785 YLVDYFKQVYSFLANGPVYVWVK-NLDLTRKQNOQKICGKFTTERDSLANVLE---QE 840  
QY 941 PQGSVLAI PASWVVDIFDLMLTPSS--CCRLYISPNKDKCPSTVNSLNCXKMSITM 998  
Db 841 RHRSTITBPLANLWDYEFMLNPQDCCRL--KKGTDEVCPPSPRRRC-ETCFQ--Q 894  
QY 999 GSVR-----PSVEQFHKYLFWFLNDRPNIKCPKGLAAYSTVLTSDQVLSRFMAY 1052  
Db 895 GSNWNNMGGFPEGKDFMEYLSIWIN-ABSDEPPLGGRAPYSTAL-VNNETSVASVFRTA 952  
QY 1053 HKPLKNSODYTEALRAARELANITADLRKVPDPAFEVPEYTTTNVYEQYLTILPBG 1112  
Db 953 HHPRLRSQKDFQAY-----SDGVRISSSPPELDMEFAYSPFYIFQYQYQTLGPLY 1001  
QY 1113 LFMLSLCLVPPFAVSCLLGLDRSGLNLSIWLVDYGFMLMIDISVAVSLINLY 1172  
Db 1002 LKLISATIIILFFISSVFL-QNIRSFLLALVWTVIYDGLMMLLGLISLNAVSLVNL 1060  
QY 1173 SAVGMSVEFVSHITSFPI--STRPTWLEBAKEATISMGSAVFPAVMTNLPGILVGL 1229  
Db 1061 IGVGIVGERCVHIVASFVVPSETKKDANSRLVSLNIGESVINGITLTKFICGVCLAF 1120  
QY 1230 AKAOGLIQIFFRPLNLTLLGLHGLVLPVILSYVG 1266  
Db 1121 AQSKIFDVYFRMFTLLIIVALLHLPLPALLSLFG 1157

RESULT 5  
T30172  
transmembrane protein patched - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T30172  
R/Goodrich, L.V.; Johnson, R.L.; Milenkovic, L.; McMahon, J.A.; Scott, M.P.







```

      765 IACYCTLCLEFGNQMABSEVCIETFSHANGSFQSYLTFRKQNNNNENINIGPILMFV 824
      905 TLGYNFSSEAGMNAICSSACNNNSFTOKIQ---YATEPEQSYLAIPASSWVDDFIDWL 961
      825 EGVKKHDPKQKQKFCCTLAGCDDNSMGNKIRSLAYEMN- KGNVYLGQDVIMIDSYLQFM 883
      962 TP-SSCCRLYISGPNKDKFC 980
      884 HPRGSCCKM-----DGKQFC 898

RESULT 8
T25600
hypothetical protein C32E8.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25600
R:Gating, S.
Submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid C32E8.
A:Reference number: Z20056
A:Accession: T25600
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-933 <GAT>
A:Cross-references: UNIPROT:P91129; UNIPARC:UPI0000611CF; EMBL:U86308; PIDN:AA842325.1;
C:Experimental source: strain Bristol N2; clone C32E8
C:Genetics:
A:Gene: C32E8.8
A:Map position: 1
A:Insertions: 145/1; 177/2; 385/1; 418/3; 833/3; 878/3

Query Match      8.7%; Score 598.5; DB 2; Length 933;
Best Local Similarity 22.9%; Pred. No. 1.2e-33;
Matches 251; Conservative 181; Mismatches 362; Indels 303; Gaps 37;

      341 FQGGTGVASWPLTIVLSVIPVALAGVFTLTDPEVMSAPNSQARSEK-AFH-- 397
      14 FROLGFLICDHPFFPFFPFLFTAMGVLHNLPLSDAVLTFTPLGASQEMERMSIHEK 73
      398 ----DOHGRP---PFRINOVLTAPNRSYRYDLSLGPKNFSGILDLLELLELQER 450
      74 WPLTDNNYIPEKRAVTSREIVYTLARN---DSNIIDPKRANAVYQDLDK-----IQTR 124
      451 LRHLQVMSPEAQRNISIQDICYAPLNDPNTSLYDCINSLLQYFQNNRTLLLTANQTL 510
      125 VRVLT----- 129
      511 GQTSQVMDKHFLYCANAPLTFKDGITALA-----LSCMADYG---APVPPFLAIGY- 559
      130 -----NGHYGVKNLCQYKNGGCPSNKGVHILSLDHNHGFNITYPYFRFSGEGYI 181
      560 -----KKGKYS-----AEALIMFSLNNP-----AGPRLAQAQLMEAFLEE 599
      182 GSSLGCVTMKGEMETDLASAKAFMYLHKFHEMSYSISG-----WE---DEL 230
      600 MRAFORRMAGMF-QVTFPAERSLEDEINRTTAEDLPATSYIVIFYLSIALGSY--SS 656
      231 GRMLTQYBEDPYISTTHSQTADLGRMDLIPRIITISITLLIVSTICSLSFIDGS 290
      657 WSRVWVDSKATLGLGVAVVLAGVAAWAMGFYSYIGIRSSLVILQVPPFLVSVGADNFI 716
      291 FSIQWVLSKPIPLISITIGVVSAGAILITGVGFLSLMGMPYN-DIVGVMPFLVAVGVDMFL 349
      717 FVLEVQRLPRRPGREVEH--IGRALGVAPSMILCSISSEALCPFLGALTTPMRAVTRAL 774
      350 MVAAN-----RTSRTHVHERMGECLADAANVSIITSSSTDVLSFVGAIITIPAVQICV 405
      775 TSGLAVIDPFLQMSAFVALLSLDSKROEASR-----LDVCCVCV----- 814
      406 YTGVAIFPAFIYQITFFAACLALAMKHBASGRNLSFLIEAVSAEKRTSLSTFORLNLGS 465

```

```

      815 -PQELPPQGBGLLGLFPQKAYAPFLMLITRGVLLLF-LALFQVSLYSKCHISVGLD 872
      466 VPDHSAASHVQKQPLTSRFFGEMVAPVLMRPVVRGIAMVFWIYLGLAS-YGCSRIKEGLE 524
      873 QELALPKDSYLLDYFLFLNRYF-EVGAPVYFV-----TTLGYNFSSE 913
      525 PVNVLVEDSYAIPHYRLLEKYFMKYGOQVQIVINNAPDLRNHTSRDRVHAMVLDFAFISGR 594
      914 A-GMAAICSSAGCANNNSFTOKIOYATEPEQ-----SYLAIPASSWVDDFIDWL 961
      585 AIGMSBV-----QFWLFEMERYQKLEVOQIIDSFFYGLHHFLASKTNNPLAEDLYN- 637
      962 TPSSCCRLYISGPNKDKFCSTVNSLNLCKMSITMGSVRPSVEQFHKLTPWFLNDRPN 1021
      638 -----GMPD----- 642
      1022 IKCPKGLAAYSTVNLITSDQVLA-SRPAVYHKPKNSQDYTEALRAARELANITADL 1080
      643 -----DDNGTMVKSFRFLLGKQDLVTMTDQDADATMSPREVAAR----- 681
      1081 RKVPGTDPAEVPEPYITTVFVYQYLTILEGLFMLSCLVPTFPAVSCLLGLDLSGLL 1140
      682 -----PEFNVTTFMIMFTQYIIIPNTVQNIITIALVNIIVIAVLEFIPQPMCS-LW 733
      1141 NLSTVMILVDYGFMAIMDISYNAVSLNLVSAVGMSEVFSHITRSFAISTKPTWLER 1200
      734 VALACASIDEGVIGYMTLGVNDALSMITITNSIGSVDSYSHIAYGVNSBEDTAAGR 793
      1201 AKERTISMGSVAPGAVAMTLPGLIVYGLAKAQLQIFPFRMLLITLLGLLGLVFLPV 1260
      794 VKALSLAGWPLSQG-AMSTIIVASVLADIPALMI-VTFPKTVVLSLGLLGLVFLPV 851
      1261 ILSYV-----GPDVNPALALEQKRAEA-----VAAVMVASCNHPGRVS 1300
      852 LLSIFPARGCIIIPSSPHGHSAQIKERQIRIAISSPDLRVAVPLRASSPISFPHLE 911
      1301 -TADNIVVNSFEFSIK 1316
      912 YTDSPFVHNRSKNSIK 928

RESULT 9
T13952
membrane protein ptc2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T13952
R:Motoyama, J.; Takabatake, T.; Takeshima, K.; Hui, C.
Nature Genet. 18, 104-106, 1998
A:Title: Ptc2, a second mouse Patched gene is co-expressed with Sonic hedgehog.
A:Reference number: Z17830; MUID:98122566; PMID:9462734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1182 <MOT>
A:Cross-references: UNIPROT:Q35595; UNIPARC:UPI000001587; EMBL:AB010833; PIDN:BAA24691
A:Experimental source: strain BALB/c3T3C
C:Genetics:
A:Gene: ptc2
C:Superfamily: Drosophila membrane protein patched
C:Keywords: transmembrane protein

Query Match      8.4%; Score 583.5; DB 2; Length 1182;
Best Local Similarity 24.2%; Pred. No. 1.8e-32;
Matches 282; Conservative 154; Mismatches 433; Indels 297; Gaps 44;

      330 SFSHTITLQ-----PFG-----WGTWVASWPLTIVLSVIPVALAGVFTBELT 376
      20 SSAPHIILAGSLQAPLMRAVAFQGLLSLQCRIQHCGKVLFLGVAGALLGLRAVAVIE 79
      377 TDPVSLMSAPNSQARSEKAFHDQHFQ-PFRINOVLTAPNRSYRYDLSLGPKNFSGI 435
      80 TDLRQLNVGVSRVSGELHTYKELGEBAAITSGMLT-----QTNHSGGNVLTPE-----A 131

```



```
Qy 436 LDLDLLELELEOEERLRLQVMSPEAQRNLSLDICV---APLNPDN-----TSLYDCC 486
Db 132 LD-----LHQAALTASKGVSLYKSGMDLNKICYKSGVPLIENGMIEMLEKLFPCV 184
Qy 487 INSLQYFQNNRTLLLTANQTLMGTSQVDKMD---HFLYCANALT-----FKDGTALA 539
Db 185 ILTPDLCFMEGAK---LQGSAYLPGRPDIQWNTLDPQQLBELGFPASLEGRELLDKA 241
Qy 540 LSCMDYGAVF-----PFLA---IGYKQKDY----- 565
Db 242 QVGOAYVGRPCLDPPDRHCPSPAPNHSQAPVNADELSCGCGCHKMKMHQBELLOG 301
Qy 566 -----EAEALMTF-----SLNNYPAGDPRLAQAUKWEAFLEEMRAFORRMAG 609
Db 302 TARDLQGLLRBAALQSTFLMSPRQLYEHFRDGYTHDIGMSEQASVWLQAMQRRFVQ 361
Qy 610 MFOVTFARSLF-----DEINRTAEDLPFATSYIV--IFLYISLALGSISSWS 658
Db 363 LAQEALPANAQQIHAFSSITLDDILRAFSE---VSTRVVGGLMLLAACVTMLRWD 417
Qy 659 RVWVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSLVILQVFPFLVSVGADNIFIV 718
Db 418 --CAOSQAGVLAGVILVALAASGLCALGITTNAATTQVLPFLALGIVDIFLFA 475
Qy 719 LEYQRLPRRPGEPREVNIGRALGRVAPSMILCSLSEALICFPLGALTMPBAVTFALTSG 778
Db 476 HAFTRAP--PDTPLPERMGECLRTGTSVALTSVNNMVAFFMALVPIPALRAFSLQAAI 533
Qy 779 AVILDFLOMSAFVALLSIDSKQOASRLDVCC-----VKRPE----- 817
Db 534 VVGCNFAAVMLVFPALISDLRRRHQRDLVLCFSSPCSAQVQWLPELGDRAVPVGI 593
Qy 818 -----LPP-----PG-----QEGG-- 826
Db 594 AHLTATVQAFTHCEASSQHVITLPPQAHLSPASPLDSELYSPGSGSTRDLSQEGRG 653
Qy 827 -----LLLGFOKAYAPFLMHTKRGVLLFLALFGVSLYSMCHISVGLDQ 873
Db 654 POACRPLCAHMTLHFAFYQFAPFLIOTRAKALVTLFGALIGSLVIGATVLOGLAL 713
Qy 874 ELALPRDSVILDFLELNRYFEVGADYVVTTLGYNFS--SEAGMAICSSACNNNSFQ 932
Db 714 TDVPRGTGEHAFLSAQLRYFSL--YEVALVTOGGFPAHSGRAL-----FDLHQ 761
Qy 933 KIQATEPEEQSYLAIPAS-----SWVDPE-----IDMLTPSSCCRLYISGPNK 976
Db 762 RSSL-----KAVLPPRATQAPRTMLHYRSMLOGIQAAPDODMASGRITCHSYRGSBD 816
Qy 977 D-----KFCPSTVNSLNCXKMSITW-----GSVPSVEQPHKYLPMFLNDRPNIKCPK 1026
Db 817 GALAYKLITQNGNAQPLDPSQLTRKLVDEKGLIP--ELFYMGTLVWVSSDPL----- 869
Qy 1027 GGLAA-----YST---SYNLSRSDQVLAIRPMAYIKPLKNSODYTEAR 1067
Db 870 -GLAASQANFYPPPEWMHDKYDTGTGENRIRPAQPLEPAOPFLHGHQKADFEALTE 928
Qy 1068 AARELAANI--TADLRVPGTDPAFBEVFPYTTNVFEQYLTLLPEGLFMLSIC--LVPTF 1124
Db 929 GARPACTEGQAGVNAVPSGSP-----LFMEQYLG--RRCFLAIVCILLVCF 976
Qy 1125 AVSCILLGLDRSGLNLISVILVDTVGFMAIMDSYNAVSLNLVSAQMSVEFVSH 1184
Db 977 LVCAILLISPMWAGIL--VLVAMMTVELFGIMGFLGIKLSAIPVILVIVASIGVEFVH 1035
Qy 1185 ITRSPAIISTKPTWLEAKAATISMSGAVRAGV--AMTVLPGILVIGLAKAQILQIFFR 1241
Db 1036 VALGFLTSHSGSRNLRAA-----SALEQTPAPVTDGAVSTLGLMLAGSNDFIIRYFV 1090
Qy 1242 LNLITLILGLHGLVLPVILSVGP 1267
Db 1091 VLTFTLILGLHGLLPLVLLTLP 1116
```

```
RESULT 10
T26746
Hypothetical protein Y39A1B.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26746
R:Wall, M.
Submitted to the EMBL Data Library, January 1998
A:Reference number: Z20258
A:Accession: T26746
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1003 <MW>
A:Cross-references: UNIPROT:Q9XXR9; UNIPARC:UP100000782PA; EMBL:AL021482; PIDs: CAA16339
A:Experimental source: clone Y39A1B
C:Genetic8:
A:Gene: CESP:Y39A1B.2
A:Map position: 3
A:Introns: 89/2; 126/3; 167/2; 194/2; 232/1; 266/3; 295/2; 323/2; 363/1; 519/3; 696/3;

Query Match
Best Local Similarity 22.8%; Pred. No. 7.3e-32;
Matches 236; Conservative 198; Mismatches 376; Indels 225; Gaps 42;

Qy 339 OFFQGWGTW--VASWPLTILVSVI PVVALAGLVTE--LTPDVELMSAPNSQARSE 393
Db 9 RFAHAFGAVSGVVAHPFPFIIPIITLAGLSTGLRRHQAFMKDELLEYTPDQAARKE 68
Qy 394 KA-----PHDQFGFPFRNQ-----VILTANRBSRYRDSILLGPKNSGLIDL 440
Db 69 ISQDLHPHINDDFVYTRRYDIRRAGYIYT-----NQEDGDLNPLVNSHMQWS 123
Qy 441 LLELELEOEERLRLQVMSPEAQRNLSLDIC-----YAPLNPNTSLYDCI 487
Db 124 IVQSITVED-----EDDRKINPSICVKFPPIPEFSKALHSLRAPNMTTBEICV 173
Qy 488 -NSLYQFQNNRTLLLTANQTLMGOT-----SQVDWKDHFLYCANAPLTFKDGITAL 540
Db 174 SNPLVEIFK-----LLVSDRSFLNRSIDEMTSLQI-----SDAIPDSGAMTHL 218
Qy 541 SCMAHYGAVFPFLAIGY---KGDYSEAEELMTFSLNNYPAGDPRLAQAUKWEAF 597
Db 219 -----LGGVTLDDDRRIGAKMMLPYALRH--SSDDEDWAEKWEVRLA 261
Qy 598 EEMRAFORRMAGMFOVTFARSLSEDEINRTAEDLPFATSYIYIFLYISLALSGSSGM 657
Db 262 DFLQYD---SPITIASMTWTETTLAEGARDBLQILHMLPFCVCSIF--TIACCVEFW 317
Qy 658 SHWVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSLVILQVFPFLVSVGADNIFIE 717
Db 318 RR-----SRPWLAIQGVISAAAMAIASAVGILLAGYGMTSVAYS--MPFIYFSGVNVFLL 372
Qy 718 VLEYQRLPRRPGEPREVNIGRALGRVAPSMILCSLSEALICFPLGALTMPBAVTFALTSG 777
Db 373 ISAMSGST--ETLEHRRKETFADAASVITVSLTDLISFVGATPPSPQMCAYAV 430
Qy 778 LAVIIDFLQMSAFVALLSIDSKQOASRLDVCCV-----KQELPPGQEG----- 826
Db 431 AAVITTYIQTLFFAAVWVYINRKEINR---HIFPHLKKQTLPEKIAQGRSPFK 486
Qy 827 --LLLGFOKAYAPFLMHTKRGVLLFLALFGVSLYSMCHISVGLDQELAPKDSYLL 884
Db 487 NTLAQOFRTTSDFLNPLVAVILVFCVYIGVASYGCTVKVLGLBNDDLPENSYOK 546
Qy 885 DYFLFLNRYF--EVGAPVYVVTTLGYNFS--EAGMAICSSACNNPSFQKIQYATEPE 942
Db 547 RFLMAAEKFFSDYGSLSL---VMMYTNLSEVDVAPRKINR-----VLEKEVELYE 592
Qy 943 OSYLAIPASSWVDLFDW-----LTPSSCCRLYISGNKDKFCPSTVNSLNCXKMS 995
Db 593 HTEFTASSDSWLRTLFAPKQAGLLITPE-----NFIYILKN--- 629
```

QY 996 ITMGVSRPVSVEQFHKKLPMFLNDRPNIKCPKGGLAAYSTSVNLTSDGVL-ASRFMAHYK 1054  
 Db 630 -----VFLSQPQF-----AKNRPVVLTEDEGHELEASRI-----658  
 QY 1055 PLK-----NSQDYTEALRAARELANITADLRKVPGRDPAFEVFPYITITVFEYOYLITP 1110  
 Db 659 PVOLRHVGSANQSRAMKLFRRLLAE--TSELQ-----IGVYADPFO-----FAQOYNALVP 706  
 QY 1111 EGLFMLSLCLVPPFAVSCLLIGDLRSGLNLISYIWIIVDPVGFMAIMDSYNAVSLIN 1170  
 Db 707 GTLSTLAVAGVAVAVASLILPEPVAS-LWVSFSIYSINIGIGFTMFVSRDLFTISMT 765  
 QY 1171 LVSAGMSVEFVSHITRSPAISTRKPTWLEPAKEATISMSGAVPAGVAMTNLPGLILVGLA 1230  
 Db 766 IVMSIGFCVDFAAHLAVNFPAKGONMDSERRNALYAVGAPILMS-ATSTIIGVSEFMA 824  
 QY 1231 KAQLQIFPRRLMLLTLLGLHGLVFLPYIS--YVGPVNPALMLEQRAEAAV- 1287  
 Db 825 ESYVFR-SFLKTMVLTLGLHGLVTLPLVLSMFYCG-----SSKAKEHIDAVD 875  
 QY 1288 -MVASCPNHPSRVST 1301  
 Db 876 QKLAQYNNPARKAS 890

RESULT 11  
 T18291  
 patched protein - zebra fish  
 C/Species: Brachydanio rerio (zebra fish)  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: T18291  
 R/Concordet, J.P.; Lewis, K.E.; Moore, J.W.; Goodrich, L.V.; Johnson, R.L.; Scott, M.P.;  
 Development 122, 2835-2846, 1996  
 A/Title: Spatial regulation of a zebrafish patched homologue reflects the roles of sonic  
 A/Reference number: Z18860; MIMD:96379744; PMID:8787757  
 A/Accession: T18291  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1220 <CON>  
 A/Cross-references: UNIPROT:Q98864; UNIPARC:UP100001328B3; EMBL:X98883; P1DN:CAA67386.1  
 C/Genetic: Pci1  
 A/Genes: Pci1  
 C/Superfamily: Drosophila membrane protein patched

Query Match 8.3%; Score 572; DB 2; Length 1220;  
 Best Local Similarity 22.4%; Pred. No. 1.2e-31; Mismatches 431; Indels 328; Gaps 43;  
 Matches 270; Conservative 174; Mismatches 431; Indels 328; Gaps 43;

QY 321 KGTSLSDKLSFTHTLLGPFQCGMTVVASWPLTIVLSYIPVVALAGLVFTELTTPV 380  
 Db 51 KGAIVGQKAPLMWRARFQALFSLGCHIQHCGKVLFIGLVGALSVGRVAALETDIE 110  
 QY 381 ELWSAPNSQARSKAFFHQHG--PFRFTNOVIIITAPNRSRYKDSLILGPKFSGILDD 439  
 Db 111 KLMVEAGSRYSKELRYTKEQGESVFTSOMLIQTPEQEG-----TNILITOE 157  
 QY 440 LLELLELQERLHLQWSPSEAQRNLSLDICVAPLNP--DNT-----SLYDCCINSL 490  
 Db 158 AL--LHLEALASAKVQVSLYKGSMDLNKICFKSGVPIIENWIERMIDKLPFCMIVTP 215  
 QY 491 LQYFQNNRTL-----LLLTANQ--TLMGQTSQVMDKHFLYCA 526  
 Db 216 LDFCWGSKLQGGSAVILPGMPDIQMNLDLKLMEBSQTSLSLEGREMIDKQOVGHAYM 275  
 QY 527 NADLTFKDGITALALSCNADYAP-----VPPPLAI--GGYKG-----561  
 Db 276 NRECLDPSDT-----DC--PHSAPNDKPMQVFNIAELQGGCHGFSKKFMHMOBELILGR 329  
 QY 562 -KD-----YSRAEALLIMTFSL-----NNYPADRPPLAQAKL-----WEAPLFE 599  
 Db 330 VKDSQNALQSAELQTFMLMSPKQLYEHFKDDDEIHIDIMNDEKATALLIESQKRFVEV 389  
 QY 600 MRAF--QRRWAGMFOVFTFAERSLDEINRTAEDLPIFATSYIVLFLYLSLALGSYSW 657

Db 390 VHGSIIPONSSSNVYAFSTT---TLNDIMKSFSDVSVYIRVAGGLMLMLAAYCTM---LRW 443  
 QY 658 SRVWDSKATLIGLGVAVVLAGVAMAMGFSSYLIGRSSVLILOVPEPLVLSVAGADNIFTF 717  
 Db 444 D--CAISQAGVLAGLVLVALSVAGLGLCSLLGLSFMNATTOVLSLALGIVDDMFL- 500  
 QY 718 VLEYQRLPRRPGEPREHVIAGLRVAPSMULCSLSEAI CFFIGALTTPMPAVTFALTS 777  
 Db 501 -LGHSFTETRSNIPFKERTGDCLRRTGTVALTSVNNMIAFMAALVPIIPALRAFSLQAA 559  
 QY 778 LAVIIDLPLQMSAFVALLSLDSKQEPASRLDVCC-----VKQOE-----817  
 Db 560 VVVVFNFMAALILFPAILSLDLHRRDKRLDILCCPYSPCSSRVIOIQBELSDANDNQ 619  
 QY 818 -----LPPRQ-----823  
 Db 620 RABATPTTGTSTTTSTHTTTVOAFTQCDAAQGHVTLIPFSQISTTPPSMVLSTPTP 679  
 QY 824 -----GEGT-----LLGFPQKAYAPFLHWITR 846  
 Db 680 TTDVPSQVFTTSSSTRDLQAQVEBPKEGECVPLPFFRMNLSSFAREKXAPFLKPEFK 729  
 QY 847 GVVLILFLALFGVSLYSMCHISVGLDQELAPKDSYLDLYFLFLNRYFVGAVYFVTL 906  
 Db 740 TVVVVVFVALLSLSTGYTMVHDGLYLTDIVPDTOGEYEFITAOQKCYFSF-YNMYLVITMD 798  
 QY 907 GYNFSEAGMNAICSSAGCNNSFTQKIQYANEFPPQSFLA-----IPASSVVDPIFW 960  
 Db 799 GPDYAR-----SORLLQHNHAFNSKTYVVKDGNHKL-RMMLHYFODV 841  
 QY 961 LTFSSCC-----RLYISGPNKDFCPSVNSLNCIKNC 993  
 Db 842 LKGLQTFPADMEAGKITDYSVRNGTEGALAVKPLIQGSKKEPFPNSQLTSRLVDG- 900  
 QY 994 MSTMGSVRPSVEQFHKTYR-WFLNDRPNIKCPKGGIA-----YST---1034  
 Db 901 ---DGLIPPEV--FYIYTLVWVNSD-----PLGYAQAQNFYPPHREMIHDKYDTJGS 948  
 QY 1035 SVNLTSDQVLSRFMAHYKPKLNSQDYTEALRAARELAN-ITADLRKVPGRDPAFEV 1093  
 Db 949 NLRIPAABPEPFAQFPFYLNGLRQASDFEALBSVTTCEBFRQOIKYTPNG-----Y 1002  
 QY 1094 PYTTNVFEYOYLITLPEGLFMLSCLVPTFAVSCLLIGDLRSGLNLISYIWIIVDPV 1153  
 Db 1003 PF-----LFWEOYIIGLHMFLLISVYLACTFLVCAILLNPMTAGVI-VFILPMVLELF 1057  
 QY 1154 GFMAIMDSYNAVSLINLVSAGMSVEFVSHITRSF--AISTKPTWLEPAKEATISMGSA 1211  
 Db 1058 GIMGLIGIKLSAIPVVLITASVIGVEFTVHIALGFLTAIGDRNT---RSAAVMBHMFAP 1114  
 QY 1212 VPAGVAMTNLPGLVILGLAKAQLQIFPRRLMLLTLLGLHGLVFLPYISVVG- 1269  
 Db 1115 VIDG-AISTLGLVLMAGSEFPIMRYFPAVALITLLIGLGLVLLPVLISLMGPABV 1173  
 QY 1270 NPA 1272  
 Db 1174 VPA 1176

RESULT 12  
 T29590  
 hypothetical protein f55f8.1 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: T29590  
 R/Gatung, S.; Scheet, P.; Kemp, K.  
 Submitted to the EMBL Data Library, November 1996  
 A/Description: The sequence of C. elegans cosmid f55f8.  
 A/Reference number: Z20647  
 A/Accession: T29590  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA

A:Residues: 1-889 <GAT>  
A:Cross-references: UNIPROT:P91346, UNIPARC:UPI0000178A20, EMBL:U80447, PIRN:AAB37812.1  
A:Experimental source: strain Bristol N2; clone F55F8  
C:Genetics:  
A:Gene: CESP:F5F8.1  
A:Map position: 1  
A:introns: 36/3; 66/3; 98/3; 149/2; 200/3; 240/2; 279/1; 358/1; 481/2; 569/3; 606/3; 697/3

Query Match 7.8%; Score 539.5; DB 2; Length 889;  
Beet Local Similarity 23.0%; Pred. No. 1.5e-29;  
Matches 223; Conservative 156; Mismatches 337; Indels 253; Gaps 35;

Dy 401 FGPFRTNOVL-----TAPNSSRYDSLLGPKNFGSILLDLLE 446  
|||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 31 FGPSYTERRIHDPMLVDGTAVAGRAVTS--REVGAAVAASGGNIIDRVSENLKT 88  
447 LOEHLRHLQWVSPEAQRNISLDICYPALPNDNTSLVDCINSILQYQNRRITLLLTAN 506  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Dd 89 MESFRN-NIVQGSNRKWSPADCLA--GPDGR--CANNDHQ----- 127  
507 QTLMGOTSQVDMKHFLYCANAPLTFKDGTALALSCHADYGAPVPFLAIGYK----- 560  
128 --LASRLHQ-----HGINITPTVRLSDKSAVYS-----ALGGVLAAGDN 167  
Dy 561 GKD-YSEKAALIMFSLNNYP----AGDRLLQAOKMEAFLEENRAFQRMAAGNQV 613  
168 GENIIVEETAWLLLYOLKFEPNEISYVSG-----LMEREFFPKMDEY-KXOAKYISI 218  
614 TETAERSLDEDINTTAEDLPFATSYVIFELY-----ISLAGSYSSWRVWDKAT 667  
219 TYFHSQTSLDELNAERLARPKFTGAFAVLTCFSYLCSVTIKSGGYIDM---VTKPI 274  
Dy 668 LGLGVAAVVLGAVMAMGFPSYLGIRSSLVILQVPPFLYSVGADNIFIPLYEQRLPRR 727  
|||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Dd 275 LSVIGVSNAGGIASAMGMITYLEIQYN-DLIAWPFVVAVGTDNMFLWASLKTRDR- 332  
728 PGEPREVHIGALGRVAPSMTLCISEAICCFGLALTMPAVRTPALTSGLAVIDELQ 787  
Dy 333 -NLKYDRIACMDADAAYSIIALTALDSFGVGITTTIPAVQFCIYTMCCALLTFAYQ 391  
788 MSAPALLSDSKQE-----ASTLDV---CCVKPDELPPPOG---- 824  
392 LTFFCALLVYTRIIEQGSHIWLRPAVYSTSPSLNKLFWLGSQPQK-PLEPGVISS 450  
Dy 825 -----EGLLGFQOKVAPFLIH-WITRGVVLLLFALGVSHTXS 863  
451 TSSVSTMTSQATSPASKIHLHCATSFFRMNYAPVLMQWI-RALAGMYLIYIGISTYG 509  
Dy 864 MCHISVGLDELAPKDSYLLDYFLFLNR-YEVGAPVYEVTT-----LGNESSE 913  
510 CTHLEKEGLEPANLVDDSYATPHYRVLEKHWGHGASHQIVASNPPDLRDVEVERINDKM 569  
Dy 914 AGNMAICGSA-GCNNFSF-----TQKIQTATE---FPREGSYLAIPASSWYDFI 958  
570 ASTFANCVYAAGDSVOQLREMOMVSEEIHKIQYDNEKPYDHAAQIYIDSOGPWADV 629  
Dy 959 DMLTPSSCCRUYISGPNKDFCPTSVNSLANCLKMKMSITWGSVRSVEQFKYLPWFAND 1018  
630 -----WGRNN 634  
Dy 1019 RPNICKPKGLAAAYSTVNLTSDGOVLARSFMAHYHKPKNSODYTEALARAEALANITA 1078  
635 -----NSERIIKTFRFMIGMRIDISTTJKQTEAINTPREIASRP-- 672  
Dy 1079 DLKRYPGDPAFEVPPYTITNVFYEQYLTILPEGFMLSLCLPFPVASCILLGLDLRSG 1138  
673 -----EOYNVTTYPLWLFDTQYALVVENPMODIIVAVACMLVISALLIQPVCSF 723  
1139 LLNLISIVMLIVDTGFWALMDISYNAVASLNLVNASGMSVEFVSHTRSFASIKKPWL 1198  
724 WV-AVTISSIDJGVGFMTLNWNDAISMKITTIINSVFSDYSAHIYAVAIKSESTTS 782  
Dy 1199 ERAKEATISMGSAVFAGVAMTNLPGLIVLGLAKQOLIOPFRRLNLITLLGLHLGVFL 1258

DB	Accession	Protein Name	Species	Accession	Protein Name	Species	Accession	Protein Name	Species		
Db	783	ARVCDALGDLHPVAQG-AMSTILAVSVLSDVPAYMI-VTFKTYFLAISTGFLHGLVFL	Caenorhabditis elegans	783	ARVCDALGDLHPVAQG-AMSTILAVSVLSDVPAYMI-VTFKTYFLAISTGFLHGLVFL	Caenorhabditis elegans	783	ARVCDALGDLHPVAQG-AMSTILAVSVLSDVPAYMI-VTFKTYFLAISTGFLHGLVFL	Caenorhabditis elegans		
Qy	1259	PVILS-YVG 1266	Caenorhabditis elegans	1259	PVILS-YVG 1266	Caenorhabditis elegans	1259	PVILS-YVG 1266	Caenorhabditis elegans		
Db	841	PLMLSVFVG 849	Caenorhabditis elegans	841	PLMLSVFVG 849	Caenorhabditis elegans	841	PLMLSVFVG 849	Caenorhabditis elegans		
RESULT 13				RESULT 13			RESULT 13				
	127969			127969			127969				
		hypochemical protein ZK675.1 - Caenorhabditis elegans				hypochemical protein ZK675.1 - Caenorhabditis elegans					
		C:Species: Caenorhabditis elegans				C:Species: Caenorhabditis elegans					
		C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Jul-2004				C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Jul-2004					
		R:Stems, M.				R:Stems, M.					
		submitted to the EMBL Data Library, November 1994				submitted to the EMBL Data Library, November 1994					
		A:Reference number: Z20448				A:Reference number: Z20448					
		A:Accession: T27969				A:Accession: T27969					
		A:Status: preliminary; translated from GB/EMBL/DBJ				A:Status: preliminary; translated from GB/EMBL/DBJ					
		A:Molecule type: DNA				A:Molecule type: DNA					
		A:Residues: 1-1405 <WTL>				A:Residues: 1-1405 <WTL>					
		A:Cross-references: UNIPROT:Q09614; UNIPARC:UPI0000083545; EMBL:Z46812; PIDN:CAA6843.1.1				A:Cross-references: UNIPROT:Q09614; UNIPARC:UPI0000083545; EMBL:Z46812; PIDN:CAA6843.1.1					
		C:Experimental source: clone ZK675				C:Experimental source: clone ZK675					
		C:Genetics:				C:Genetics:					
		A:Gene: CESP:ZK675.1				A:Gene: CESP:ZK675.1					
		A:Map position: 2				A:Map position: 2					
		A:Introns: 31/1; 59/3; 100/3; 244/2; 485/1; 950/2; 1298/2; 1335/3				A:Introns: 31/1; 59/3; 100/3; 244/2; 485/1; 950/2; 1298/2; 1335/3					
		C:Superfamily: Drosophila membrane protein patched				C:Superfamily: Drosophila membrane protein patched					
		Query Match				Query Match					
		Best Local Similarity 7.4%; Score 514; DB 2; Length 1405;				Best Local Similarity 7.4%; Score 514; DB 2; Length 1405;					
		Matches 267; Conservative 190; Mismatches 412; Indels 452; Gaps 52;				Matches 267; Conservative 190; Mismatches 412; Indels 452; Gaps 52;					
Qy	321	KGTSLSDKLSFSTHTLGLQFFQGWGWV--ASWPLTILVSVYPVVALAAGLVFELTTD	378	Qy	321	KGTSLSDKLSFSTHTLGLQFFQGWGWV--ASWPLTILVSVYPVVALAAGLVFELTTD	378	Qy	321	KGTSLSDKLSFSTHTLGLQFFQGWGWV--ASWPLTILVSVYPVVALAAGLVFELTTD	378
Db	100	RGKATGNRYALYSRLIQKLFALGNTVHRNAMS-I LAVSMIFAVC-CYGLQYHIETD	157	Db	100	RGKATGNRYALYSRLIQKLFALGNTVHRNAMS-I LAVSMIFAVC-CYGLQYHIETD	157	Db	100	RGKATGNRYALYSRLIQKLFALGNTVHRNAMS-I LAVSMIFAVC-CYGLQYHIETD	157
Qy	379	PVELMSA-----PMSQARSEKAFIDQHGPFF-RNV-----QVLTLPNRRS	419	Qy	379	PVELMSA-----PMSQARSEKAFIDQHGPFF-RNV-----QVLTLPNRRS	419	Qy	379	PVELMSA-----PMSQARSEKAFIDQHGPFF-RNV-----QVLTLPNRRS	419
Db	158	IVKLVNVAAGGRLDEELNPLPNIKEMRNVTGGS-GPELPRENGAGGYQVLDIQPEYEG	215	Db	158	IVKLVNVAAGGRLDEELNPLPNIKEMRNVTGGS-GPELPRENGAGGYQVLDIQPEYEG	215	Db	158	IVKLVNVAAGGRLDEELNPLPNIKEMRNVTGGS-GPELPRENGAGGYQVLDIQPEYEG	215
Qy	420	YRYSILGPKKFGSILDLLELLELLEQLERLHLOWSPEQRNI-----SDIDICY	472	Qy	420	YRYSILGPKKFGSILDLLELLELLEQLERLHLOWSPEQRNI-----SDIDICY	472	Qy	420	YRYSILGPKKFGSILDLLELLELLEQLERLHLOWSPEQRNI-----SDIDICY	472
Db	216	--QDALAAGP-----LLKHYEIMKHILASFVSVHGVDWMSLSDICF	253	Db	216	--QDALAAGP-----LLKHYEIMKHILASFVSVHGVDWMSLSDICF	253	Db	216	--QDALAAGP-----LLKHYEIMKHILASFVSVHGVDWMSLSDICF	253
Qy	473	APLNPD-----NLSLYD-----CC-INSLLQYFQNNRTL-----LLI	503	Qy	473	APLNPD-----NLSLYD-----CC-INSLLQYFQNNRTL-----LLI	503	Qy	473	APLNPD-----NLSLYD-----CC-INSLLQYFQNNRTL-----LLI	503
Db	254	KPAPPVSVAADSAAISGLVDIKIVPCIWITPIDCFWESKALGPHPSLPKSLGPLYGMLL	313	Db	254	KPAPPVSVAADSAAISGLVDIKIVPCIWITPIDCFWESKALGPHPSLPKSLGPLYGMLL	313	Db	254	KPAPPVSVAADSAAISGLVDIKIVPCIWITPIDCFWESKALGPHPSLPKSLGPLYGMLL	313
Qy	504	TANQTLMGQTSQVWPKD-----HFLYCANAPLTF-----	532	Qy	504	TANQTLMGQTSQVWPKD-----HFLYCANAPLTF-----	532	Qy	504	TANQTLMGQTSQVWPKD-----HFLYCANAPLTF-----	532
Db	314	SS---ISDGMIRMSDPDPIAVIDEIHRSEFLSGHYTFPEERAGVSHGYMDRPPCIDPLDP	369	Db	314	SS---ISDGMIRMSDPDPIAVIDEIHRSEFLSGHYTFPEERAGVSHGYMDRPPCIDPLDP	369	Db	314	SS---ISDGMIRMSDPDPIAVIDEIHRSEFLSGHYTFPEERAGVSHGYMDRPPCIDPLDP	369
Qy	533	-----KDGTALALSCADYAGAPVPPFLAIGGYGKDYSEAE									

[illegible]

Query Match	6.9%; Score 473.5; Db 2; Length 1015;
Best Local Similarity	19.24; Pred. No. 7.5e-25;
Matches	226; Conservative 186; Mismatches 417; Indels 347; Gaps 38;
QY	289 IILCSVAV---VYILLGFRVAPARAKSKMVDPKGTSLSDKLSFSTHTLLGQFQGWG 345
DB	11 LKTSIFBAGKNVLLALIAHRF-PMRCRIPDLKK---LEKVAWYTKHLVDY---- 61
QY	346 TWVASMPLTLIVSLVFPVVALAAGLV-TEILT-----DPV-----ELMSA 385
DB	62 -----YVPLIAVILLATCSGCFWIKETLLDARKLTPVSAIPMKERQVPSLMPV 114
QY	386 PMSQARSEKAFHDOHFGPFRTNQVILLAPRSSYRDSLLGKMFSGILLDLLEL 445
DB	115 KSYEFLEPRT-----FQMRNYLVVHGREGNGT-----YPNLEGSYLANDIA 158
QY	446 ELQERL-----RHLOWSPEAQRNLSLODICTAPLPNPTSLYD 484
DB	159 KIEBELVTNVSFPREALEDNPALMRKRHLNETEIEENRNTI-----TFKD 205
QY	485 CCINSLIQPQNNRTLLLTANQTLMOGTSQVPMKHFIYCANAPLPFKOSTALASGMA 544
DB	206 VQNMWYDQCKREKNVIBLKRRLN-----KGISVTPQVN 243
QY	545 DYGAVPPEPLAIG---YKGDVSEAEALIMTFSLNNYPAGDPRLAOAKME---EAFLE 598
DB	244 QEGPPIYAFYIGVVDTPENDTIKYAAPMRIMYFLKDDDEQEOU--AKFPEDTAEKFR 301
QY	599 EMRAFQRRMAGMPQVTFATERSLDEINRTAEADLPITATSUYIIPYISIALGSYS--- 655
DB	302 ETYADHPRI---QCHKHSRIYDQGLTRANNRKLKPFNVNTIANLVLEFPA---YSKMX 353
QY	656 -----SWSRWYVD--SKATLGGVAVVIGAVMAAGFFSYGIRSSLYIIOVPPV 706
DB	354 YFRMDHWP-LHIDWMSKRPMLALGVLSSVIALISGILLMPGMEFAEITL-LAPFLV 411
QY	707 LSVGADNIFITVLEY---QRLPRRQEPREVNHIGALGRVAPBMLCSLSEALCFPLGA 762
DB	412 LSIQVDDMFILVAAHMTKEMKYPERSPKVMKRMIEAMSESAVAIFITSPFDVLSFGAGT 471
QY	763 LTPMPAVRTFALYTSGLVAILDFLLQMAFVALLSBDRQOASLNDVCCV----- 813
DB	472 ITDIIAVOGFCAMTAAOCFFTEFLYQITFPALMVISAKAQMGSNNSCMPCTIAGDIYTE 531
QY	814 -----KQELPP 820
DB	532 DGLQPNLKKKKSKRKDAKAEKKQDQAKNDKMEIBENAEAKESYSYDSSPPSQIHI 591
QY	821 PGQSEGILLGFQKAYAPFLHMWITRGVLLFLFALFGVSLYSMCHISVGHDOELAPKD 880
DB	592 PVKSRGAMGHFRDIYPPMLNMKTKLFMPFTPIIYIGISYVGI CWNEGQSDYDKLLHS 651
QY	881 SYLLDYELFLNRYEYVGAPVYFVTLTGYNFSSGAMNAICSSAGCNCNFSPTOKIQVATEF 940
DB	652 DPIVE---ALGRREL-----FHGGDQLEIA--- 674
QY	941 PEGSYLAIASSWDDFDIMLTLPSSCCRLYISGNKDKFCPSTYNSLNLCKMCSITMGS 10000
DB	675 -----IONCENITIAE 685
QY	1001 VRPSVEQFHKLPMFLNDRPIKPKKG-----LAAYSVNLTSQ-----GQ 10433
DB	686 SRDRIE-----LVAQEFENISYSLGKGTSFMWRREYKXYSNLGSLYANDRESWITGV 738
QY	1044 VLARFMAVHK-----PLKNSOQY-TEALPAAR-----ELAAVITADLRKVP 10844
DB	739 YEMSOLEFAFYCLMGOQFWMANESYDTELMSYFRIGVHRLSPFTDLVL-ITSEIAGVA 797
QY	1085 GTDPAFEVFPRTINVFYEQYLLTIPBGLFMSLCLVTPFAVSCILLGLDLSRGLNLS 11444
DB	798 DRHEDNINVTYQGSRAIADQNLVLLSITNTDITLAMCMCFVALIIFIPNP-CALFITFA 856
QY	1145 IVMILVDTVGFMALMDSIYNAVSLINIVSAVGMSVEFVSHITRSFALS-TKPTWLERAKE 12031

Db 857 WTITDGVGFSLMSVKLDPISMTITISIGSIFSAHTHGFVSNDSNLSAFRCVD 916  
Qy 1204 ATISMGSAAVAGVAMTNCPLGLVLGLAKAQLIOIFFEFLNLTLLGLHGLVFLPILLS 1263  
Db 917 AMEKLAAMPVHG-SLSTIGVFLVARIDSYVMVLVFETKISIVL-IIGAHMLMLPILLS 974  
Qy 1264 YVGP-----DVNPALALQKRAEEAVAAVMVASCP 1293  
Db 975 WCPIVTERLSDASKASDRRRKLEKKNKNSVYAINLP 1010  
RESULT 15  
S06119  
membrane protein patched - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: S06119; A33468  
R:Nakano, Y.; Guerrero, I.; Hidalgo, A.; Taylor, A.; Whittle, J.R.S.; Ingham, P.W.  
Nature 341, 508-513, 1989  
A:Title: A protein with several possible membrane-spanning domains encoded by the Drosoph  
A:Reference number: S06119; MUID:90015164; PMID:2797178  
A:Accession: S06119  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1299 <NAK>  
A:Cross-references: UNIPROT:P18502; UNIPARC:UPI000014EA18; GB:X17558; NID:g8389; PIDN:CA  
R:Hooper, J.E.; Scott, M.P.  
Cell 59, 751-765, 1989  
A:Title: The Drosophila patched gene encodes a putative membrane protein required for se  
A:Reference number: A33468; MUID:9008658; PMID:2582494  
A:Accession: A33468  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: mRNA  
A:Residues: 1-110, 'R', 112-273, 'G', 275-331, 'R', 333-635, 'P', 637-861, 'DVF', 878, 'Y', 880-1299  
A:Cross-references: UNIPARC:UPI000016BCA3; GB:M26418; GB:M28099; NID:g552097; PID:g55209  
C:Genetics:  
A:Gene: FlyBase:ptc  
A:Cross-references: FlyBase:FBgn0003892  
A:Map position: 2 4AD3-D4  
C:Superfamily: Drosophila membrane protein patched  
C:Keywords: glycoprotein; transmembrane protein  
F:74-92/Domain: transmembrane #status predicted <TM01>  
F:427-448/Domain: transmembrane #status predicted <TM02>  
F:456-503/Domain: transmembrane #status predicted <TM03>  
F:529-555/Domain: transmembrane #status predicted <TM04>  
F:557-585/Domain: transmembrane #status predicted <TM05>  
F:677-699/Domain: transmembrane #status predicted <TM06>  
F:967-1017/Domain: transmembrane #status predicted <TM07>  
F:1019-1047/Domain: transmembrane #status predicted <TM08>  
F:1061-1086/Domain: transmembrane #status predicted <TM09>  
F:1093-1121/Domain: transmembrane #status predicted <TM10>  
F:142,298,335,388,807,861,1194,1271/Binding site: carbohydrate (Asn) (covalent) #status  
Query Match 6.7%; Score 464.5; DB 1; Length 1299;  
Best Local Similarity 20.4%; Pred. No. 4,4e-24;  
Matches 245; Conservative 174; Mismatches 472; Indels 311; Gaps 41;  
Qy 321 KETSLSDKLSFSTHTLLGPFQGMGTWVSWPLTILVSVIPVLAALGLVTELTDDV 380  
Db 43 KKKARSSRAIYLRYSVQSHLETLGSSVGKAGKVLFAILVLSFTCVLKAQHSKVH 102  
Qy 381 ELSMSPNSQARSEKAFNDHFG--PFRITNOVILTA---PNRSSYRDSLGLGPKNPSGIL 436  
Db 103 QLMWEGGGLLEALATYQKTIGEDBSATHQLLIQTHDPAASVLAHQAALL-----A 153  
Qy 437 DLDLLELLELRNLHNVSPPEAQRNLSLODICYAPLNPDNTSLY-----DCC 486  
Db 154 HLEVLVAKATVAKVNLVDTLW-----GLRDMCMSPSTSPFEGIYUIEQILRLHLP 204  
Qy 487 INSLLOYFN-----NRTLLLLTAN-----OTLMGQSQVDMKQH 521  
Db 205 ITPPLDCEFWESQQLGPESAAVVIPLGNRLMTTLNPAVQMYMKQMSBEKISFDETV 264

Qy 522 FLYCANAPLTFKDGATLALSCM-----ADGAPVPPFLAIG--GYK 560  
Db 265 EQYMRRAIA--SGYMKRPPCLNPLNCPDTPAPKNSTOPDVA-----ILSGCYGA 317  
Qy 561 GKDYSAEALIMTFSLNYPADPRLAQA----- 589  
Db 318 AKHNMVPEELLVGGAKRN--RSGHLRKAQALQSVOVQMTKEKMYDQMYKVNHGVTQE 376  
Qy 590 -----KLMEBAF--LEEMRAFORRMAGMFOVTFPABRSLEDEINRTAEDLPFATSY 640  
Db 377 KAAEVLNMQNRFNSREVEQLLRKQRIANDVDIYVSSAALDDLAKFSHSAISVIGV 436  
Qy 641 IYIFLYISLALGSYSNSRWVWDSKATGLGCAVAVLCAVMAKGFSLIGRSLVLQ 700  
Db 437 AVTVLY--AFCTLLRW--RDPVRCQSSVGVAGVILMCFSTYAGLGLSALLGIVFNASTQ 492  
Qy 701 VVPELVLSGADNIFIFLVEYQRLPRRGEFREYHIGALRVAPSMILGISEALICFL 760  
Db 493 VVPELALGLGVDHIFMLTAAVAESNRR--EQTKL---ILKKVGSILFSACTAGSFPA 546  
Qy 761 GALTMPAVRTPALTSGLAVILDFLOKSAFVALISDSKQASRLDV-CCCV-----K 814  
Db 547 AAFIVPALIKVFCLOALVMCSNLAALLVPRAMISLDRRTGRADIFCCCFPMWKEQ 606  
Qy 815 PQELPP-----PGQEGEGL-----LIGPF 832  
Db 607 PKVAPPPVPLNNNGRGARHPKSCNNNRVALPAQNPLEGRADIPGSSHSIASFSIATFA 666  
Qy 833 OKAYAPPELLMHTTRVNVLLFLALFGVSLYEMCHISVGLDEMLPKXSILDYFLPNR 892  
Db 667 FQHYTFPFRSWVKFLTWGFLDALISLVASTRLODGLDIDILVPKDSNEHKFLDAQTR 726  
Qy 893 YEEVAPVYFVTLTGYNFSSSEKGMNAICSSAGCNFSFTOKIQVATEPEQSYLAIP--- 949  
Db 727 LF--GFYEMVAVTQG-----NFEVPTQQLLRQD--HDSFVAVPHVI 764  
Qy 950 -----ASSVVDIFDWLTPSSCCRLYISGBNKKFCPSYVNSLNCILKNC----- 993  
Db 765 KNDNGCLPDPFWLLLFSEWL-----GNLQKIDEEYRDGLRTKECMFPNASSDA 812  
Qy 994 -----MSITMGSVRSVQ-----FHKYL--PWFLDGRPNIKCPKGG 1028  
Db 813 ILAYVLIVQTHGVNDPVRKVELVLTNRVNSDGIINORAFVNYLSAM-----ATMSSPTEL 868  
Qy 1029 LAA--YSTSVNLTPSDGVLAARFMAVHKLKNSOD-----YTEALPRAAR 1070  
Db 869 LRANGIRRRANGASGGLKYPERQVFNQ--NEYDLKIPKSLPVYAAQMPFYLHGLDPTS 926  
Qy 1071 ELAANITADLRKVPCTDPAFEV--FPYTTTVNVEYQVLTILPEGLFMLSCLVPTFAVSC 1128  
Db 927 QI-KTLIGHIRDLISVYEGFGILPNYPSGIPFIWFQWYT--LRSSLAMILACVLALVAV 984  
Qy 1129 LLLGLDASGLNLSTYMWLVDTVGFPAALMDISNAVSLNIVSAVQMSVEFVSHTRS 1188  
Db 985 SLLLSISWAAVVLVLSIASLAQIFGAATTLGIKISALPAVILLSVSMKCF--NVLIS 1042  
Qy 1189 FAISTKPTWLEBAKATISMGSAAVPAVAMTNLPGLVGLAKAQLIOIFFEFLNLTLL 1248  
Db 1043 LGPMTSVGNRRQKRVOLSNQMSLGRPLVHGMLTSGAVFPLTSPPEFVRLHNCMLLVLC 1102  
Qy 1249 LGLHGLVFLPYLISYVGPVNPALALQKRAEEAVAAVMVASCPNHSRVSVDNIYVN 1308  
Db 1103 VGACNSLIVFPIILSMVPEAE-LVPLLE-----HDIRISTSPPLPV 1143  
Qy 1309 HS 1310  
Db 1144 SS 1145

Search completed: March 22, 2006, 22:56:20  
Job time : 63 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 22:47:27 ; Search time 265 Seconds  
(without alignments)  
3546.278 Million cell updates/sec

Title: US-10-736-769-4  
Perfect score: 6909  
Sequence: 1 MBRAGRGMLMLALLRLAQ.....GSIKAGATSNFLPNNRGPF 1332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6872.5	99.5	1359	NPCL1_HUMAN	O9UHC9 homo sapien
2	5421.5	78.5	1331	NPCL1_RAT	O6C3U3 rattus norv
3	5407	78.3	1333	NPCL1_MOUSE	O6C3U4 mus musculu
4	3179	46.0	1132	Q4T749_TETNG	Q4T749 tetradon n
5	2413	34.9	1277	Q9J1G3_CRIGR	Q9J1G3 cricetulus
6	2412	34.9	1277	NPCL1_PIG	P56941 sus scrofa
7	2405	34.8	1276	Q8MKD8_FELCA	Q8MKD8 felis silve
8	2404.5	34.8	1276	Q9GKS2_CANFA	Q9GKS2 canis famli
9	2402.5	34.8	1276	NPCL1_HUMAN	O15118 homo sapien
10	2402	34.8	1289	Q59GR1_HUMAN	Q59GR1 homo sapien
11	2400	34.7	1277	Q7TMD4_MOUSE	Q7TMD4 mus musculu
12	2399	34.7	1276	Q9N000_FELCA	Q9N0Q0 felis silve
13	2395	34.7	1276	Q8W149_FELCA	Q8W149 felis silve
14	2385	34.5	1278	NPCL1_MOUSE	O35604 mus musculu
15	2367.5	34.3	1277	Q9GLC9_BOVIN	O9GLC9 bos taurus
16	2361	34.2	1286	Q9TTF5_RABIT	O9TTF5 oryctolagus
17	2338	33.7	1209	Q4RWY5_TETNG	Q4RWY5 tetradon n
18	2262	33.7	1287	Q9V124_DROME	Q9V124 drosophila
19	2256	33.7	1287	Q9U5M1_DROME	Q9U5M1 drosophila
20	2253	33.6	1287	Q7YU59_DROME	Q7YU59 drosophila
21	2205	31.9	1291	Q7Q409_ANOGA	Q7Q409 anopheles g
22	1864.5	27.0	1223	Q9VRC9_DROME	Q9VRC9 drosophila
23	1760	25.5	1003	Q7PS03_ANOGA	Q7PS03 anopheles g
24	1725	25.0	1075	Q5INK7_MAGGR	Q5INK7 magnaporthe
25	1712.5	24.8	1361	Q7XUB7_ORYSA	Q7XUB7 oryza sativ
26	1700.5	24.6	1271	Q5BBG1_EMENI	Q5BBG1 aspergillus
27	1659.5	24.0	1330	Q5XKS9_CRYNE	Q5XKS9 cryptococcu
28	1658.5	24.0	1330	Q5SSD4_CRYNE	Q5SSD4 cryptococcu
29	1643	23.8	1273	Q4MMG5_ASPTU	Q4MMG5 aspergillus
30	1614.5	23.4	1295	Q410K4_GIBZE	Q410K4 gibberella
31	1549	22.4	1264	Q6BT03_DEBHA	Q6BT03 debaryomyce

## ALIGNMENTS

RESULT 1	NPCL1_HUMAN	STANDARD	PRT	1359 AA.
AC	Q9UHC9; Q6R3Q4; Q9UHC8;			
DT	13-SEP-2005 (Rel. 48, Created)			
DT	13-SEP-2005 (Rel. 48, Last sequence update)			
DT	13-SEP-2005 (Rel. 48, Last annotation update)			
DE	Niemann-Pick CL-like protein 1 precursor.			
GN	Name=NPCL1;			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Eukaryota; Metazoa; Chordata; Euteleostomi; Primates; Catarrhini; Homidae;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RM	1;			
RP	NCBIOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 2 AND 3), AND TISSUE SPECIFICITY.			
RP	SPECIFICITY.			
RX	MEDLINE=20247253; PubMed=10783261; DOI=10.1006/geno.2000.6151;			
RA	Davies J.P., Levy B., Ioannou Y.A.,			
RT	"Evidence for a Niemann-Pick C (NPC) gene family: identification and			
RT	characterization of NPCL1."			
RL	Genomics-65:137-145(2000).			
RM	[2]			
RP	NCBIOTIDE SEQUENCE [MRNA] (ISOFORM 2), AND TISSUE SPECIFICITY.			
RX	PubMed=14976318; DOI=10.1126/science.1093131;			
RA	Altman S.W., Davis H.R., Jr., Zhu L.-U., Yao X., Hoos L.M.,			
RT	Terzloff G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,			
RT	Murgolo N., Gratianno M.P.;			
RT	"Niemann-Pick CL like 1 protein is critical for intestinal cholesterol			
RT	absorption."			
RL	science 303:1201-1204(2004).			
RM	[3]			
RP	TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.			
RX	PubMed=15671032; DOI=10.1074/jbc.M409110200;			
RA	Davies J.P., Scott C., Oishi K., Liapis A., Ioannou Y.A.;			
RT	"Inactivation of NPCL1 causes multiple lipid transport defects and			
RT	protects against diet-induced hypercholesterolemia."			
RL	J. Biol. Chem. 280:12710-12720(2005).			
RM	[4]			
RP	INDUCTION.			
RX	PubMed=15504518; DOI=10.1194/jlr.M400400-JLR200;			
RA	van der Veen J.N., Krull J.K., Havinga R., Baller J.F.W., Chimini G.,			
RT	Lestavel S., Stels R., Groot P.H.E., Groen A.K., Kuipers F.;			
RT	"Reduced cholesterol absorption upon PPARdelta activation coincides			
RT	with decreased intestinal expression of NPC1L1."			
RL	J. Lipid Res. 46:526-534(2005).			
RM	[5]			
RP	FUNCTION.			
RX	PubMed=15528087; DOI=10.1073/pnas.0500269102;			
RA	Garcia-Calvo M., Lisnock J., Bull H.G., Hanes B.E., Burnett D.A.,			
RA	Braun M.P., Crona J.H., Davis H.R., Jr., Dean D.C., Detmers P.A.,			
RA	Gratianno M.P., Hughes M., Macintyre D.B., Ogawa A., O'Neill K.A.,			
RA	Iyer S.P.N., Shevell D.E., Smith M.M., Tang Y.S., Makarewicz A.M.,			
RA	Ujjainwalla F., Altman S.W., Chapman K.T., Thornberry N.A.;			

32	1527.5	22.1	1275	2	Q9SHN9_ARATH	Q9SHN9 arabidopsis
33	1511.5	21.9	1342	2	Q9TVK6_DICDI	Q9TVK6 dictyosteli
34	1508	21.8	1239	2	Q6CBAL_YARLI	Q6CBAL yarrowia li
35	1500	21.7	1162	2	Q7RWLV_NEUROA	Q7RWLV neurospora
36	1461.5	21.2	1256	2	Q59ZV0_CANAL	Q59ZV0 candida alb
37	1452.5	21.0	1397	2	Q551C5_DICDI	Q551C5 dictyosteli
38	1437.5	20.8	1489	2	Q4PEB3_USRMA	Q4PEB3 uscilago ma
39	1394	20.2	1055	2	Q9SVF0_ARATH	Q9SVF0 arabidopsis
40	1386.5	20.1	620	2	Q4H344_CIOIN	Q4H344 clostridia
41	1359	19.7	1382	2	Q618V4_CABBR	Q618V4 caenorhabdi
42	1341.5	19.4	1383	1	NPCL1_CABEL	NPCL127 caenorhabdi
43	1329	19.2	1170	2	Q122D0_YEAST	Q122D0 saccharomyc
44	1323	19.1	1339	2	Q5ORB6_ENTHI	Q5ORB6 entamoeba h
45	1312.5	19.0	1178	2	Q750G1_ASHGO	Q750G1 ashbya goes



```

RT "The target of ezetimibe is Niemann-Pick C1-Like 1 (NPC1L1).";
RL Proc. Natl. Acad.Sci. U.S.A. 102:8132-8137(2005).
RN [6]
RP VARIANTS LEU-55 AND ASN-1233.
RX PubMed15679830; DOI=10.1111/.1399-0004.2004.00388.x;
RA Wang J., Williams C.M., Hegele R.A.;
RT "Compound heterozygosity for two non-synonymous polymorphisms in NPC1L1 in a non-responder to ezetimibe.";
RL Clin. Genet. 67:175-177(2005).
CC -1- FUNCTION: Play a major role in cholesterol homeostasis. Is critical for the uptake of cholesterol across the plasma membrane of the intestinal enterocyte. Is the direct molecular target of ezetimibe, a drug that inhibits cholesterol absorption. Lack of activity leads to multiple lipid transport defects. The protein may have a function in the transport of multiple lipids and their homeostasis, and may play a critical role in regulating lipid metabolism.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation of brush border membranes from proximal enterocytes suggests considerable association with the apical membrane fraction. Exists as a predominantly cell surface membrane expressed protein (By similarity). According to Ref.3 localizes in a subcellular vesicular compartment rich in RAB5.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9UHC9-1; Sequence=Displayed;
CC Name=2; Synonyms=NPC1L1DELTAE15;
CC IsoId=Q9UHC9-2; Sequence=VSP_015314;
CC Name=3; Synonyms=NPC1L1;
CC IsoId=Q9UHC9-3; Sequence=VSP_015312, VSP_015313;
CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed in liver. Also expressed in small intestine, pancreas, kidney, lung, pancreas, spleen, heart, gall bladder, brain, testis, stomach and muscle.
CC -1- INDUCTION: Expression is decreased in Caco-2 cells upon PPARd activation.
CC -1- PTM: Highly glycosylated (By similarity).
CC -1- POLYMORPHISM: Variations in NPC1L1 gene could be associated with nonresponse to ezetimibe treatment.
CC -1- SIMILARITY: Belongs to the patched family.
CC -1- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
CC -----
CC EMBL, AF192522; AAF20396.1; -, mRNA.
CC EMBL, AF192523; AAF20397.1; -, mRNA.
CC EMBL, AY515256; AAS6939.1; -, mRNA.
CC EMBL, AY437865; AAR97886.1; -, mRNA.
CC Ensembl: ENSG0000015520; Homo sapiens.
CC HGNC; HGNC:7898; NPC1L1.
CC MIM; 608010; -.
CC GO; GO:0015918; P:sterol transport; TAS.
CC InterPro; IPR004765; NP_C type.
CC InterPro; IPR003392; Patched.
CC InterPro; IPR000731; SSD 5TM.
CC Pfam; PF02460; Patched; 1.
CC TIGRFAMs; TIGR00917; ZAO60601; 1.
CC PROSITE; PS50156; SSD; 1.
CC Alternative splicing; Cholesterol metabolism; Glycoprotein; Lipid metabolism; Polymorphism; Signal; Steroid metabolism; Transmembrane.
KW SIGNAL.
FT 1 21 Potential.
FT CHAIN 22 1359 Niemann-Pick C1-like protein 1.
FT TOPO_DOM 22 284 Extracellular (Potential).
FT TRANSMEM 285 305 1 (Potential).
FT TOPO_DOM 306 351 Cytoplasmic (Potential).
FT TRANSMEM 352 372 2 (Potential).
FT TOPO_DOM 373 632 Extracellular (Potential).
FT TRANSMEM 633 653 3 (Potential).

```

FT	TORO DOM	654	666	Cytoplasmic (Potential)	4 (Potential)	
FT	TRANSMEM	667	687	Extracellular (Potential) <td></td> <td></td>		
FT	TORO DOM	668	696	Extracellular (Potential) <td></td> <td></td>		
FT	TRANSMEM	697	717	5 (Potential) <td></td> <td></td>		
FT	TORO DOM	718	742	Cytoplasmic (Potential) <td></td> <td></td>		
FT	TRANSMEM	743	763	6 (Potential) <td></td> <td></td>		
FT	TORO DOM	764	776	Extracellular (Potential) <td></td> <td></td>		
FT	TRANSMEM	777	797	7 (Potential) <td></td> <td></td>		
FT	TORO DOM	798	846	Cytoplasmic (Potential) <td></td> <td></td>		
FT	TRANSMEM	847	867	8 (Potential) <td></td> <td></td>		
FT	TORO DOM	868	882	Extracellular (Potential) <td></td> <td></td>		
FT	TRANSMEM	883	903	9 (Potential) <td></td> <td></td>		
FT	TORO DOM	904	1139	Cytoplasmic (Potential) <td></td> <td></td>		
FT	TRANSMEM	1140	1160	10 (Potential) <td></td> <td></td>		
FT	TORO DOM	1161	1168	Extracellular (Potential) <td></td> <td></td>		
FT	TRANSMEM	1169	1189	11 (Potential) <td></td> <td></td>		
FT	TORO DOM	1190	1191	Cytoplasmic (Potential) <td></td> <td></td>		
FT	TRANSMEM	1192	1212	12 (Potential) <td></td> <td></td>		
FT	TORO DOM	1213	1236	Extracellular (Potential) <td></td> <td></td>		
FT	TRANSMEM	1237	1257	13 (Potential) <td></td> <td></td>		
FT	TORO DOM	1258	1359	Cytoplasmic (Potential) <td></td> <td></td>		
FT	DOMAIN	632	797	SSD.		
FT	COMBIAS	500	503	Poly-Leu.		
FT	CARBOHYD	54	54	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	132	132	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	138	138	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	244	244	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	416	416	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	431	431	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	464	464	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	479	479	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	497	497	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	506	506	N-linked (GlcNAc . .)	(Potential)	
FT	CARBOHYD	626	626	N-linked (GlcNAc . .)	(Potential)	
FT	VASPLIC	723	724	RL -> GP (In isoform 3)		
FT	VASPLIC			/FtId=VSP 015312		
FT	VASPLIC	725	1359	Missing (In isoform 3)		
FT	VASPLIC			/FtId=VSP 015313		
FT	VASPLIC	1046	1072	Missing (In isoform 2)		
FT	VARIANT	55	55	V -> L (nonresponse to ezetimibe treatment)		
FT	VARIANT			/FtId=VAR 023369		
FT	VARIANT	1233	1233	I -> N (nonresponse to ezetimibe treatment)		
FT	VARIANT			/FtId=VAR 023370		
SEQ	SEQUENCE	1359 AA;	148698 MW;	3225D53D93B768B6 CRC64;		
Query Match		99.5%;	Score 6872.5;	DB 1;	Length 1359;	
Best Local Similarity		97.9%;	Pred. No. 0;			
Matches 1330;	Conservative	0;	Mismatches	2;	Indels	27;
					Gaps	1;
QY	1	MAEAGRGWMLALLRLA	JOSEPYTTTHHOPGYCAFYDECGKNPELSSGLMTLSNVSCLSN	60		
DB	1	MAEAGLRWMLALLRLA	JOSEPYTTTHHOPGYCAFDECGKNPELSSGLMTLSNVSCLSN	60		
QY	61	TPARKITGDHLLLOKICPR	LYTGPNTQACCSAQOLVSLBASLSTTKALLTRCPACSDNF	120		
DB	61	TPARKITGDHLLLOKICPR	LYTGPNTQACCSAQOLVSLBASLSTTKALLTRCPACSDNF	120		
QY	121	VNLHCNHTCSPNQSLF	INVTVAOLGAGQLPAVVAVEAFYOHSPFAEQSYDSCSRVVPAA	180		
DB	121	VNLHCNHTCSPNQSLF	INVTVAOLGAGQLPAVVAVEAFYOHSPFAEQSYDSCSRVVPAA	180		
QY	181	ATLAVGTCGYGYSALCNA	QRWLNFGQDPTNGCLAPLDTTFHLLRPGQAVGSGIOPLNCGV	240		
DB	181	ATLAVGTCGYGYSALCNA	QRWLNFGQDPTNGCLAPLDTTFHLLRPGQAVGSGIOPLNCGV	240		
QY	241	ARCNSQGDVAVTSCOD	CAASCPAIPAPALDSTFYVGMPGSLVLIILICSVPAAVTI	300		
DB	241	ARCNSQGDVAVTSCOD	CAASCPAIPAPALDSTFYVGMPGSLVLIILICSVPAAVTI	300		
QY	301	LIVAFRVA	PARDKRNVDPKKGTSLSDLSFSTHTLLGQPFQMGWTWASWPLTTLVLSV	360		



```

Db      ||||| 301 LTVGFVAPARDKSKVNDPKQSTLSDKLSFSFTHLLGQFQGWGVMVSWPLTITLVLV 360
Qy      ||||| 361 IPVVALAAGLVFTELTTDPVELMSAPNSQARSKAHFDHGFPPFTNQVILTAPNRSSY 420
Db      ||||| 361 IPVVALAAGLVFTELTTDPVELMSAPNSQARSKAHFDHGFPPFTNQVILTAPNRSSY 420
Qy      ||||| 421 RYDSLLGPKNFSGILLDLLELLELLOERLRLQVMSPEAQNISLQDI CYAPLNDPT 480
Db      ||||| 421 RYDSLLGPKNFSGILLDLLELLELLOERLRLQVMSPEAQNISLQDI CYAPLNDPT 480
Qy      ||||| 481 SLYDCINSLQFQNNRRTLLLTANQTLMGQTSQVDMKHFLYCANAPLTFDQGRALL 540
Db      ||||| 481 SLYDCINSLQFQNNRRTLLLTANQTLMGQTSQVDMKHFLYCANAPLTFDQGRALL 540
Qy      ||||| 541 SCMAADGAPVFPPLAIGYKGDYSEAEALIMFESLNNYPAGPRLAQAALMEAELEEM 600
Db      ||||| 541 SCMAADGAPVFPPLAIGYKGDYSEAEALIMFESLNNYPAGPRLAQAALMEAELEEM 600
Qy      ||||| 601 RAEQRMAGMFOYTFPABERSLEDEINRTAEDLPFATSYIVIFLYISLALGSSWSRV 660
Db      ||||| 601 RAEQRMAGMFOYTFPABERSLEDEINRTAEDLPFATSYIVIFLYISLALGSSWSRV 660
Qy      ||||| 661 MDSKATLIGYAVVLGAVMAAMGFSTYIGIRSSLVILQVPLVLSGADNIFTVLE 720
Db      ||||| 661 MDSKATLIGYAVVLGAVMAAMGFSTYIGIRSSLVILQVPLVLSGADNIFTVLE 720
Qy      ||||| 721 YORLPRGPBPVRVHIGRALGRVAPMLCSSEALCFELGALTMPVARTPLTSGLAV 780
Db      ||||| 721 YORLPRGPBPVRVHIGRALGRVAPMLCSSEALCFELGALTMPVARTPLTSGLAV 780
Qy      ||||| 781 IIDFLQMSAFVALSLDSKROASRLDYCCVCPQELPPQEGEGILLGFPOKAVAPFL 840
Db      ||||| 781 IIDFLQMSAFVALSLDSKROASRLDYCCVCPQELPPQEGEGILLGFPOKAVAPFL 840
Qy      ||||| 841 LHMITRGVULLFLALFGVSLYSMCHISVGLDOELALPKDSYLDYFLFNRYEYVAGPY 900
Db      ||||| 841 LHMITRGVULLFLALFGVSLYSMCHISVGLDOELALPKDSYLDYFLFNRYEYVAGPY 900
Qy      ||||| 901 YPFTTLGYNPSSAGNNALCSSAGCNPNFSTOKIOVATEPEBOSYALTPASSVNDPFIW 960
Db      ||||| 901 YPFTTLGYNPSSAGNNALCSSAGCNPNFSTOKIOVATEPEBOSYALTPASSVNDPFIW 960
Qy      ||||| 961 LTPSSCCRLYISGPNKDCEPSTVNSLNCJKNCMSITMGSVRPSVEQFHXYLPWFLNDR 1020
Db      ||||| 961 LTPSSCCRLYISGPNKDCEPSTVNSLNCJKNCMSITMGSVRPSVEQFHXYLPWFLNDR 1020
Qy      ||||| 1021 NIKCPKGLAAVSTSVNLTSQGVLT-----ASRFMAVH 1053
Db      ||||| 1021 NIKCPKGLAAVSTSVNLTSQGVLT-----ASRFMAVH 1053
Qy      ||||| 1054 KPLKNSQDYTEALRAARELANITADLRKVRGDPAPAEVPPYITTVFPEQYITIPBGL 1113
Db      ||||| 1054 KPLKNSQDYTEALRAARELANITADLRKVRGDPAPAEVPPYITTVFPEQYITIPBGL 1113
Qy      ||||| 1081 KPLKNSQDYTEALRAARELANITADLRKVRGDPAPAEVPPYITTVFPEQYITIPBGL 1140
Db      ||||| 1081 KPLKNSQDYTEALRAARELANITADLRKVRGDPAPAEVPPYITTVFPEQYITIPBGL 1140
Qy      ||||| 1114 FMLSCLVPTFAVSCILLGLDARSGLNLSIYMIIVDTVGFALMDISNAVSLNLSV 1173
Db      ||||| 1114 FMLSCLVPTFAVSCILLGLDARSGLNLSIYMIIVDTVGFALMDISNAVSLNLSV 1173
Qy      ||||| 1141 FMLSCLVPTFAVSCILLGLDARSGLNLSIYMIIVDTVGFALMDISNAVSLNLSV 1200
Db      ||||| 1141 FMLSCLVPTFAVSCILLGLDARSGLNLSIYMIIVDTVGFALMDISNAVSLNLSV 1200
Qy      ||||| 1174 AVGMSVEFVSHITRSPALSTKPTWLBRAKEATISMGSAVAGVAMTNLPGLIVLGLAKQ 1233
Db      ||||| 1174 AVGMSVEFVSHITRSPALSTKPTWLBRAKEATISMGSAVAGVAMTNLPGLIVLGLAKQ 1233
Qy      ||||| 1201 AVGMSVEFVSHITRSPALSTKPTWLBRAKEATISMGSAVAGVAMTNLPGLIVLGLAKQ 1260
Db      ||||| 1201 AVGMSVEFVSHITRSPALSTKPTWLBRAKEATISMGSAVAGVAMTNLPGLIVLGLAKQ 1260
Qy      ||||| 1234 LIOIFFFRNLTLTLGLHGLVFLPYIILSYVGPDPVNPALALQKABEAVAAVMAVASC 1293
Db      ||||| 1234 LIOIFFFRNLTLTLGLHGLVFLPYIILSYVGPDPVNPALALQKABEAVAAVMAVASC 1293
Qy      ||||| 1261 LIOIFFFRNLTLTLGLHGLVFLPYIILSYVGPDPVNPALALQKABEAVAAVMAVASC 1320
Db      ||||| 1261 LIOIFFFRNLTLTLGLHGLVFLPYIILSYVGPDPVNPALALQKABEAVAAVMAVASC 1320
Qy      ||||| 1294 NHRSRVSTADNIVYNSFEGSISGAGAISNPLFNNRGOF 1332
Db      ||||| 1294 NHRSRVSTADNIVYNSFEGSISGAGAISNPLFNNRGOF 1332
Qy      ||||| 1321 NHRSRVSTADNIVYNSFEGSISGAGAISNPLFNNRGOF 1359
Db      ||||| 1321 NHRSRVSTADNIVYNSFEGSISGAGAISNPLFNNRGOF 1359

```

RESULT 2

```

NPCL1_RAT
ID NPCL1_RAT STANDARD; PRT; 1331 AA.
AC 061303;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Niemann-Pick C1-like protein 1 precursor.
GN Name=Npc1l1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND SUBCELLULAR
RP LOCATION.
RC STRAIN=Sprague-Dawley;
RX PubMed=14976318; DOI=10.1126/science.1093131;
RA Altman S.W., Davis H.R. Jr., Zhu L.-O., Yao X., Hoos L.M.,
RA Tetziolf G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,
RA Murgolo N., Graziano M.P.;
RT "Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol
RT absorption.";
RL Science 303:1201-1204 (2004).
RN [2]
RP SUBCELLULAR LOCATION, AND CHARACTERIZATION.
RX PubMed=15777641; DOI=10.1016/j.bbag.2004.12.021;
RA Iyer S.P.N., Yao X., Crona J.H., Hoos L.M., Tetziolf G.,
RA Davis H.R. Jr., Graziano M.P., Altman S.W.;
RT "Characterization of the putative native and recombinant rat sterol
RT transporter Niemann-Pick C1 like 1 (NPCL1) protein.";
RL Biochim. Biophys. Acta 1722:282-292 (2005).
RN [3]
RP FUNCTION.
RX PubMed=15928087; DOI=10.1073/pnas.0500269102;
RA Garcia-Cairo M., Lisnock J., Bull H.G., Hawes B.E., Burnett D.A.,
RA Braun M.P., Crona J.H., Davis H.R. Jr., Dean D.C., Demers P.A.,
RA Graziano M.P., Hughes M., Macintyre D.E., Ogawa A., O'Neill K.A.,
RA Iyer S.P.N., Shevell D.E., Smith M.M., Tang Y.S., Makarewicz A.M.,
RA Ujjainwalia F., Altman S.W., Chapman K.T., Thornberry N.A.;
RT "The target of ezetimibe is Niemann-Pick C1-like 1 (NPCL1).";
RL Proc. Natl. Acad. Sci. U.S.A. 102:8132-8137 (2005).
CC -!- FUNCTION: Play a major role in cholesterol homeostasis. Is
CC critical for the uptake of cholesterol across the plasma membrane
CC of the intestinal enterocyte. Is the direct molecular target of
CC ezetimibe, a drug that inhibits cholesterol absorption (by
CC similarity). Lack of activity leads to multiple lipid transport
CC defects. The protein may have a function in the transport of
CC multiple lipids and their homeostasis.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation
CC of brush border membranes from proximal enterocytes suggests
CC considerable association with the apical membrane fraction. Exists
CC as a predominantly cell surface membrane expressed protein.
CC -!- TISSUE SPECIFICITY: Small intestine showed the highest level of
CC expression. Expression in other tissue including gall bladder,
CC liver, testis, and stomach is also observed. Along the duodenum-
CC ileum axis, the levels vary in different segments of intestine,
CC with peak expression in the proximal jejunum. Protein expression
CC is confined to the enterocyte. Discrete localization to the
CC epithelial layer bordering the luminal space along the crypt-
CC villus axis. Protein expression in the enterocyte is observed
CC closest to the luminal space. Expression in enterocytes from the
CC proximal (jejunum) but not in the distal (ileum) region.
CC -!- PTM: Highly glycosylated.
CC -!- SIMILARITY: Belongs to the patched family.
CC -!- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.
CC This Swiss-Prot entry is copyrigh. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.

```

DR EMBL: AY437867; AAR97888.1; -; mRNA.  
DR RGD: 1303135; Npc11.  
DR InterPro: IPR004765; NP\_C\_type.  
DR InterPro: IPR003392; Patched.  
DR InterPro: IPR000731; SPD\_5TM.  
DR Pfam: PF02460; Patched; 1.  
DR TIGRFAMs: TIGR00917; 2A060601; 1.  
DR PROSITE: PS50156; SPD\_1.  
KW Cholesterol metabolism; Glycoprotein; Lipid metabolism; Signal;  
KW Steroid metabolism; Transmembrane.  
FT SIGNAL 1 20 Potential.  
FT CHAIN 1 1331 Niemann-Pick C1-like protein 1.  
FT TOPO\_DOM 21 282 Extracellular (Potential).  
FT TRANSMEM 283 303 1 (Potential).  
FT TOPO\_DOM 304 352 Cytoplasmic (Potential).  
FT TRANSMEM 353 373 2 (Potential).  
FT TOPO\_DOM 374 632 Extracellular (Potential).  
FT TRANSMEM 633 653 3 (Potential).  
FT TOPO\_DOM 654 665 Cytoplasmic (Potential).  
FT TRANSMEM 666 686 4 (Potential).  
FT TOPO\_DOM 687 696 Extracellular (Potential).  
FT TRANSMEM 697 717 5 (Potential).  
FT TOPO\_DOM 718 742 Cytoplasmic (Potential).  
FT TRANSMEM 743 763 6 (Potential).  
FT TOPO\_DOM 764 776 Extracellular (Potential).  
FT TRANSMEM 777 797 7 (Potential).  
FT TOPO\_DOM 798 846 Cytoplasmic (Potential).  
FT TRANSMEM 847 867 8 (Potential).  
FT TOPO\_DOM 868 1113 Extracellular (Potential).  
FT TRANSMEM 1114 1134 9 (Potential).  
FT TOPO\_DOM 1135 1142 Cytoplasmic (Potential).  
FT TRANSMEM 1143 1163 10 (Potential).  
FT TOPO\_DOM 1164 1165 Extracellular (Potential).  
FT TRANSMEM 1166 1187 11 (Potential).  
FT TOPO\_DOM 1187 1206 Cytoplasmic (Potential).  
FT TRANSMEM 1207 1237 12 (Potential).  
FT TOPO\_DOM 1238 1242 Extracellular (Potential).  
FT TRANSMEM 1243 1263 13 (Potential).  
FT TOPO\_DOM 1264 1331 Cytoplasmic (Potential).  
FT DOMAIN 53 797 SSD.  
FT CARBOHYD 53 53 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 85 85 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 138 138 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 244 244 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 416 416 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 431 431 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 464 464 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 479 479 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 497 497 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 506 506 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 566 566 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 606 606 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 626 626 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 909 909 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 917 917 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 996 996 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1038 1038 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1076 1076 N-linked (GlcNAc... ) (Potential).  
SQ SEQUENCE 1331 AA; 146415 MW; 2E10EF2E3A337F70 CRC64;

Query Match 78.5%; Score 5421.5; DB 1; Length 1331;  
Beet Local Similarity 77.8%; Pred. No. 0;  
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAEAGIRGWLIALRLAQLSEPYTTHIQGYCAFYDECKNPBELSGIMLTSLNVSCLSN 60  
DB 1 MAAMWL-GWLIALRLAQLSAQGLYTPKHAGVCTFEBCGKNPELSGGLTSLNVSCLSN 59  
QY 61 TPARKITGDHLILLOKICPRLTYGPNQ-ACCSAKOLVLEASLSITKALLTRCPACSDN 119  
DB 60 TPARYHTGHEHLALORICPRLYNGPNTTACCSKOLLSLESSMSITKALLTRCPACSDN 119  
QY 120 FVNLCHNTGSPDLSFINTVTRVAVOAGAGOLPAVVAVEAFQHSFPEQSYDSCSRVAV 179  
DB 120 FVSLCHNTGSPDLSFINTVTRVVERGAGEPPAVVAVEAFQHSRAEKAYSCQVRLPA 179

DB 120 FVSLCHNTGSPDLSFINTVTRVVERGAGEPPAVVAVEAFQHSRAEKAYSCQVRLPA 179  
QY 180 AATLAVGTMGVYSGALCNAQRMLNFOGDTNGALAPLDTITHLLEPGAVSGIQPLNEG 239  
DB 180 AASLAVSGMGVYSGALCNAQRMLNFOGDTNGALAPLDTITHLLEPGALPDGIQPLNGK 239  
QY 240 VARNESGQDDVATSCDDCAASCALARPQALBSTFLGCMPSGLVLIILLCSPAVAVT 299  
DB 240 IAPCHESQGDSDAVSCDDCAASCVCVIPPPLARPSFWMGPGLALIIITFAFVFLS 299  
QY 300 ILVGFRAVPARADSKQWDPKKGTSLSDKLSFSTHTLLGQFFQGGTVAWSPLTIVLS 359  
DB 300 AVLVALRVSRNKNKAGCPQBAEPPLPKHLKLSPTITIGRFQWGTAVASMPLTIVLS 359  
QY 360 VLPVVALAAGVETLTTPDVELWSAPNSQARSEKAFDHPGPFRTNQVLTAPNRS 419  
DB 360 FIVVALAAGVETLTTPDVELWSAPNSQARSEKAFDHPGPFRTNQVLTAPNRS 419  
QY 420 YRYDLSLLGPKNFGSLDLDELLELQERLRHQVWSPERAKRISLQDICAPLAPDN 479  
DB 420 KYDLSLLGSKNFGSLDLDELLELQERLRHQVWSPERAKRISLQDICAPLAPDN 479  
QY 480 TSLVCCINSLLQYQNNRTLLLTANQTLNGQTSQVDMKHFLYCANAPLTFKDGTA 539  
DB 480 TSLSDCCNSLLQYQNNRTLLMTNQTNGQTSQVDMKHFLYCANAPLTFKDGTS 539  
QY 540 LSCMADYAPVFPFLAIGYKGDYSEAEALIMFSLNNYPAGDBRLAQAKMEAPLE 599  
DB 540 LSCMADYAPVFPFLAIGYKGDYSEAEALIMFSLNNYPAGDBRLAQAKMEAPLE 599  
QY 540 LSCMADYAPVFPFLAIGYKGDYSEAEALIMFSLNNYPAGDBRLAQAKMEAPLE 599  
DB 540 LSCMADYAPVFPFLAIGYKGDYSEAEALIMFSLNNYPAGDBRLAQAKMEAPLE 599  
QY 600 MRAPRBAAGPQVTFMERLEDEINTTMDPIFAFTSYIVFLYLSLALGSYSMSR 659  
DB 600 MESFRANSDRFQVAFSERSLEDEINTTQDDLPVFVASYIIVFLYLSLALGSYSRCSR 659  
QY 660 VVVDKATLIGGVAVVVGAVMAAMFPFSYIGIRSSLYLQVFPVLVSGADNIFIVL 719  
DB 660 VVVDKATLIGGVAVVVGAVMAAMFPFSYIGIRSSLYLQVFPVLVSGADNIFIVL 719  
QY 660 VVVDKATLIGGVAVVVGAVMAAMFPFSYIGIRSSLYLQVFPVLVSGADNIFIVL 719  
DB 660 VVVDKATLIGGVAVVVGAVMAAMFPFSYIGIRSSLYLQVFPVLVSGADNIFIVL 719  
QY 720 EYQRLPRRGPBEVHIGRALGRVAPSMULCSLSEALCFGLGALTMPAVRTFALTSL 779  
DB 720 EYQRLPRRGPBEVHIGRALGRVAPSMULCSLSEALCFGLGALTMPAVRTFALTSL 779  
QY 720 EYQRLPRRGPBEVHIGRALGRVAPSMULCSLSEALCFGLGALTMPAVRTFALTSL 779  
DB 720 EYQRLPRRGPBEVHIGRALGRVAPSMULCSLSEALCFGLGALTMPAVRTFALTSL 779  
QY 780 VILDLTQMSAPVALLSDSKQEAERLUDVCCVKKPQGLPPPGQEGILLGFFQKAVAPF 839  
DB 780 VILDLTQMSAPVALLSDSKQEAERLUDVCCVKKPQGLPPPGQEGILLGFFQKAVAPF 839  
QY 780 IILDLTQMSAPVALLSDSKQEAERLUDVCCVKKPQGLPPPGQEGILLGFFQKAVAPF 839  
DB 780 IILDLTQMSAPVALLSDSKQEAERLUDVCCVKKPQGLPPPGQEGILLGFFQKAVAPF 839  
QY 840 ILHMTTRGVLLFLALFGVSLYSWCHTSVGLDQELAPKDSYLDLYFLNRYEYVAP 899  
DB 840 ILHMTTRGVLLFLALFGVSLYSWCHTSVGLDQELAPKDSYLDLYFLNRYEYVAP 899  
QY 840 ILHMTTRGVLLFLALFGVSLYSWCHTSVGLDQELAPKDSYLDLYFLNRYEYVAP 899  
DB 840 ILHMTTRGVLLFLALFGVSLYSWCHTSVGLDQELAPKDSYLDLYFLNRYEYVAP 899  
QY 900 VYFVTTLLGYNFSSAAGNNAICSSAGCNPFSETKIQVATBEPQSYLAIPASSWVD 959  
DB 900 VYFVTTLLGYNFSSAAGNNAICSSAGCNPFSETKIQVATBEPQSYLAIPASSWVD 959  
QY 900 VYFVTTLLGYNFSSAAGNNAICSSAGCNPFSETKIQVATBEPQSYLAIPASSWVD 959  
DB 900 VYFVTTLLGYNFSSAAGNNAICSSAGCNPFSETKIQVATBEPQSYLAIPASSWVD 959  
QY 960 WITP-SSECRUYISGPNQDKCPSTVNSLCLKXCMSTITMSVAPSVQGFHKYLPWF 1018  
DB 960 WITP-SSECRUYISGPNQDKCPSTVNSLCLKXCMSTITMSVAPSVQGFHKYLPWF 1018  
QY 960 WITP-SSECRUYISGPNQDKCPSTVNSLCLKXCMSTITMSVAPSVQGFHKYLPWF 1018  
DB 960 WITP-SSECRUYISGPNQDKCPSTVNSLCLKXCMSTITMSVAPSVQGFHKYLPWF 1018  
QY 1019 PENICPKPGGLAAYSTVNLSDGOVLASRFMAVYKPLKNSODYTEALRAARELANITA 1078  
DB 1019 PENICPKPGGLAAYSTVNLSDGOVLASRFMAVYKPLKNSODYTEALRAARELANITA 1078  
QY 1079 DLKRVPGTDPAFEVPPYITTVFQYQYLTILPEGLFMLSCLVPTFAVSCLLGLDRSG 1138  
DB 1080 DLKRVPGTDPAFEVPPYITTVFQYQYLTILPEGLFMLSCLVPTFAVSCLLGLDRSG 1138  
QY 1139 LNLILSIWILVDYTGPMALDISYNAVSLINIVSAVMSVEFVSHITRSPASTKPMWL 1198  
DB 1140 LNLILSIWILVDYTGPMALDISYNAVSLINIVSAVMSVEFVSHITRSPASTKPMWL 1198  
QY 1199 LNLILSIWILVDYTGPMALDISYNAVSLINIVSAVMSVEFVSHITRSPASTKPMWL 1199  
DB 1199 LNLILSIWILVDYTGPMALDISYNAVSLINIVSAVMSVEFVSHITRSPASTKPMWL 1199  
QY 1259 BRAKATISMGSAVPAVGAATNLPGIILVGLAKAQLOIQFFFRNLNLTLLGLHGLVFL 1258  
DB 1200 BRAKATISMGSAVPAVGAATNLPGIILVGLAKAQLOIQFFFRNLNLTLLGLHGLVFL 1259



	Matches	1028	Conservative	126	Mismatches	174	Indels	4	Gaps	4
QY	4	AGLRGWMILLALLRLA	GEPTTTHIQPGYCAFYDEBCGNKPELSLMTLSNVSCLSINTPA	63						
Db	3	AAWGQMLLMALLINSAG	QELGYTPTHKAQFCTFYEEBCGNKPELSGTLSTLSNISCLSNTPA	62						
QY	64	RKIRGDHLLLOKICPL	YLTGPN-TQACCAKCOLVSLASISITKALLTRCPACSDPNVN	122						
Db	63	RHVYGDHLLIQRVCPE	RLYNPNBNDTYACCSKTQVLSLSSISITKALLTRCPACSENFVS	122						
QY	123	LHCNHTCSPNQSLF	INTVRVAQLGAGQIPAAVAAEAFQYGHFAEQSYSCSRVAVPAAT	182						
Db	123	IHCNHTCSPDQSLF	INTVRVVGQDQGLPAAVAAEAFQGRFAEAYSCSRVAIIPAAAS	182						
QY	183	LAVGTCGCVYSALC	NAQRMLNPGDTNGLAPLDITFHLLEPGQAVSGSIQPINEGVAR	242						
Db	183	LAVGSMCGVYSALC	NAQRMLNPGQDGTNGLAPLDITFHLLEPGQALDKCKPKPDGKITP	242						
QY	243	CNBSGQDDVATCS	CODCAACCPALARPAALDSTFPLGMPBSLVLIILLCVPAVITLL	302						
Db	243	CNBSGQEDSAACS	CODCAACCPVIRPPALRPPSFYMGMPMMLIILITFVAVFLSLVL	302						
QY	303	VGFRVAPARDSK	QWDPKKGTSLSDIKLSFSFHTLLGCFQFQMGVMAWMPITLIVLSYIP	362						
Db	303	YTLRVASNRNKKTK	AGSOEARLPKKRFSFHTYVGRFESMGTRVAMPLITVALSTIV	362						
QY	363	VVALAAGLVTEL	TTDPVELWSAPNSQARSEKAFPHDQGFPPRTNQVILTPAQRSSRY	422						
Db	363	VIALSVGLTFTEL	TPDPEVLSAPKQARKKAFHDBHGFPPRINQIFVTAQRSSRYKY	422						
QY	423	DSLILGPKNFG	SIGLIDLILBLLEQLERHLQVMSPEAQRNISLQDI CYAPLPNDTSL	482						
Db	423	DSLILGPKNFG	SIGLISLDLLOELLEQERLRLHQQVMSHQAQRNISLQDI CYAPLPNHTSL	482						
QY	483	YDCCINSLLOQF	QONNRLLLLLTAAQTLMGOTSQVDVMDKHPLFCYANAPLTFQDGTALALSC	542						
Db	483	TDCCVNSLQIF	QONNHTLLLLLTAAQTLMGOTSQVDMKHPLFCYANAPLTFYQDGTALALSC	542						
QY	543	MADYGA	PVPPELAIIGYKQKDYSEAEALIMTFSLNNYPADGPRLAQAQLMEAEFLAEKRA	602						
Db	543	IADYGAPVPEFL	AVGQGYQTDYSEAEALIIFFSINNYPADDPRAHAAQLMEAEFLAEKMQS	602						
QY	603	FORRMAGNFQ	YTFMERLEDEINRTMEDLPFATSTIVFLYISLALSGYSKSRMV	662						
Db	603	FORSTADFOJA	FAFSERLEDEINRTTIQDLPVFALISYLIVFLYISLALSGYSKSRVAV	662						
QY	663	DSKATILG	AGVAVVYGVAAAMAGFSYSGISSLVTLQVPEFVLSVAGADNIPIFVLEYQ	722						
Db	663	DSKATILG	DGVAVVGVAAMAGFYSYSGVBSLVTLQVPEFLVAVAGADNIPIFVLEYQ	722						
QY	723	RLPRRPGEPRE	VHIGRALGRVAPSWMLCSLSBAICFFLGALTPMPAVTPALITGLAVIL	782						
Db	723	RLPRRPGEPRE	HEHIGRTIGSVAPSWMLCSLSBAICFFLGALTIMPAVTPALITGLALIF	782						
QY	783	DFLLQMSAFV	ALLSDSKRQEARLIDVCCYKPOELPPPGQEGEGLIGFQKAYAPFLH	842						
Db	783	DFLLQMTAFV	ALLSDSKRQEARPDVCCFSSRLPPKQKEGILLCCFFKRIYTPFLH	842						
QY	843	WITGCVVLLFL	ALPGVSLVMCHSYGLDQELALPKQSYLLDVFLPLNRFPVGA	902						
Db	843	RFIRPVVLLFL	VPYFGANVLMCNISVGLDDDLAPKQSYLLDVFLPLNRFLVGGPPVYF	902						
QY	903	VYTLGYNFSSE	AGNAICSSAGCUNNFSTQIKQYATEPEQSYALIPASSWVDFIDULT	962						
Db	903	DTTSGYNFST	EAQNAICSSAGCESFSLQIKQYASEFPNOSYVAILAASSWVDDFIDULT	962						
QY	963	P-SGCGRLYIS	GPNDKFCPSTVNSLNCIKCMCSITMGSVAPSEVQFHKYLPWELNDRN	1021						
Db	963	PSSSCCR	IYTRGPHDEFCPSTDTSFNCLCKCMNRITLGPVAPPTGQFHKYLPWFLNDTPN	1022						
QY	1022	IKCPHGGI	AAVSTSVNLTSQGVASLRMAHHPKAKNSQDTYALRARARELANIYADLR	1081						
Db	1023	IRCPHGGI	AAVARTSVNLSSDQILASQFMAHHPKAKNSQDPTTEALRARARLANIYADLR	1082						

QY	1082	KVPEPTDAFAEVPPTTINNVYEQYLTLPGLPGLMELSLCIVPTPAVSCLLGLDLRSGLLN	1141
DB	1083	KVPEPTDENFEVFPPTTISNVFYQYLTLPGLPGITFLALCFPTPTVVCYLLGLDLRSGLLN	1142
QY	1142	LSLTVMLTVDTVGSMALMDISYNAVSLINLVSAVNGSVSEVSHITSRFSFAISTKPTMLERA	1201
DB	1143	LSLSTMLVDTDTIGLMAWMGSIYNAVSLINLVTVAVNGSVSEVSHITSRFSFAISTKPTMLERA	1202
QY	1202	KEATISMGSAVPAQVANTNLPGLIVIGLAKAQLIQIFPFLNLITLLGLHGLVLPVLT	1261
DB	1203	KDATIFMGSAVPAQVANTNFPGLIILGFQAQQLIQIFPFLNLITLLGLHGLVLPVLT	1262
QY	1262	LSYVGPVNVNALMLEQRAREAVAAVVAWSCPMHPSRVSSTADNIYVNHSPSGS-1KAGA	1320
DB	1263	LSYVGPVNVNALMLEQRAREAVAAVVAWSCPMHPSRVSSTADNIYVNHSPSGS-1KAGA	1321
QY	1321	ISNPLPNNRGOF 1332	
DB	1322	ASSSLPKSDQKF 1333	
RESULT 4			
ID	Q4T749	TETNG PRELIMINARY; PRT; 1132 AA.	
AC	Q4T749		
DT	13-SEP-2005	(TREMblrel_31, Created)	
DT	13-SEP-2005	(TREMblrel_31, Last sequence update)	
DT	13-SEP-2005	(TREMblrel_31, Last annotation update)	
DE	Chromosome undetermined SCAF8317, whole genome shotgun sequence.		
DE	(Fragment).		
CN	ORFNames=GSTENG0005930001;		
OS	Tetradodon nigroviridis (Green puffer).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomoptera; Acanthopterygii; Percormorpha; Tetraodontiformes;		
OC	Tetraodontidae; Tetraodontidae; Tetradodon.		
CC	NCBI_TaxID=99883;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Jatillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,		
RA	Mauceli E., Bouneau L., Fischer C., Ozouf-Coetzac C., Bernot A.,		
RA	Nicoud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Seguirin B.,		
RA	Dasilva S., Salenobac M., Levy M., Boudet N., Castellano S.,		
RA	Anthonard V., Jubin C., Castell J.V., Katinka M., Vacherie B.,		
RA	Bismont C., Skalli Z., Cartolico L., Poulsen J., De Bernardis V.,		
RA	Craud C., Duprat S., Broctier P., Couvanceau J.P., Gouzy J.,		
RA	Parra G., Lardier G., Chapple C., McKernan K.J., McSwan P., Bosak S.,		
RA	Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,		
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,		
RA	Lander V., Schachter V., Quetier F., Saurin W., Scarpelli C.,		
RA	Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;		
RT	"Genome duplication in the teleost fish Tetradodon nigroviridis reveals		
RT	the early vertebrate proto-karyotype.";		
RL	Nature 431:946-957(2004).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RG	Genoscope, Whitehead Institute Centre for Genome Research;		
RL	Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.		
CC	-1- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; CAAB01008317; CAF91283.1; -; Genomic_DNA.		
FT	NON TER		
SO	SEQUENCE 1132 AA; 125923 MW; 4283CF4CA0BDBE97 CRC64;		
Query Match 46.0%; Score 3179; DB 2; Length 1132;			
Best Local Similarity 52.0%; Pred. No. 4.3e-216;			
Matches 641; Conservative 167; Mismatches 264; Indels 160; Gaps 11;			
QY	87	TOACCSAKQVLSLEASISTKALLTRPACSDNFVNLHCHNTGSPNGLINTNRVAGL-	145
DB	1	TVACCSINQSLSEASLASKAVILRPSCADNFAHHCITTSSPDQTKIKYKTVNVT	60

QY 146 GAGOL--PAVVAVEAFYCHFAEQSYDSCSRVRVPAATLAVGTMGVYGSALCNQORWL 203  
 DB 61 GPDNITREVEGVAAVNLSTNFPADASFQSCKNVRIPATGGLAGTMOGRGSHCNQORRY 120  
 QY 204 NFGDGTGNGLAPLDTITFHLLEPG--QAVSGIQPLNEGVACNCSGDDVATFCSQCDCA 261  
 DB 121 DFGDSSNGLAPLDTIDFHLIKPGDTEDEPGIIVPYSGNALKNCTETTPSGKACSCQDCOE 180  
 QY 262 SGPARAPALDSTFYLQMGPSLVL--IIILC-SVPAVVTIILNGR-----RV 307  
 DB 181 SCPEKTPPEPLPPGPQLGMDGFLIITILLCILFALFVYTLRRNNKKGLKKKO 240  
 QY 308 AAPARDKSNVDPKKGSLSDKLSFTHTLLGQFQMGCTVVAWPLTIIIVLSIPIVVALA 367  
 DB 241 QKSHVDTQVVISPSVTCADKNSQLAQAALSSFGQWGVMAFTVTLVLLSAILVYVLS 300  
 QY 368 AGLVTELTITDPVEIWSAPNSQARSEKAFHDQFGFPFTNOVILTAPNRSYRDSLL 427  
 DB 301 AGIKNIETLTDPEIWSAPNSRQEKAFHDTHFDFFFTNOILITAPKEGMYDSLIF 360  
 QY 428 GKPNSSGILDLLELLELORLRLQVWSPBAQNISLQDIQVAPLMDNTSLYDCCI 487  
 DB 361 GQHNISGIIISKDILILELEOKOIQNIEFWSKQDINTATLKQVCFAPLNPNSASTDCAV 420  
 QY 488 NSLLQYFQNNRTLLLTANQTLMGQTSQYDMKQHFYCANAPLTFQDGTALALSCMADYG 547  
 DB 421 NSLPQYFQNSLENINAKNMTELGVTKEDWMDHLITCNSPLSFQDITDLQMSCADYG 480  
 QY 548 APVFPFLAIGYKKGKDYSEAEALIMTFSLNYPAGDPLAQAQKWEAEFLBEMRAFQRM 607  
 DB 481 APVIFPLAVGVYQ-----MAMEMAEFLKIQVEQKNP 513  
 QY 608 AGMFOVTFAAESLEDEINRTTAEDEPIATSYIVILYITSLAGSYSSMSRM-VDSKA 666  
 DB 514 STNFTLAYAERSLEDEINRTTAEDEPIMISIAVIFLYIAVALGFTSMKRLVVDSKF 573  
 QY 667 TGLGVAVAVLGVMAAMGFYSYLGRSSLVILQVVPFVLSVGAANITFVLEYQRLR 726  
 DB 574 LVGLGILVASCNVLSLGFCSKVGIPSSLIIVQVPLVLAAGDNITFVLEYQRDVR 633  
 QY 727 RGEPEPEVHIGRALGRVAPSMILCSLEAICPELGAITMPAVRTFALTSGLAVIIDL 786  
 DB 634 RGEKREBQIIGRLGNVAPSMILCSLEAICPEFGALSTMPAVKSPALVAVALIDFAL 693  
 QY 787 QMSAFVALISDSKQOEARLDVCCVQOELPPQOGSGLIGFQKXAYAPPLHMTIR 846  
 DB 694 QMTAFVALISDORODSNRCBLCCIKVSRKPTKPNEGFLLPFMKKYAPALINRYSR 753  
 QY 847 GVALFLFALFGVSLVSMCHISVGLDOELALPKDSYLLDYPLFANKYFEVGAUVYFVTL 906  
 DB 754 IIVMFVFIIMLCGSIITLTHVAVGLDQELAMPKDSIMLKIFELVYKFEVGAUVYFVTR 813  
 QY 907 GYNFSSEAGMNAICSSAGCNFSPTOKIOYATEFPBOSYLAIPASSWVDFTMLTPSSC 966  
 DB 814 GYNFVSVMNAVCSVSGCDPYSTQKIQYATEYPELQQLDU----- 855  
 QY 967 CELYISGPKKDXFCBSTVNSLNCIKCMSTMGSVRPSVBQFHXYLPWFLNDRPKTKCX 1026  
 DB 856 -----IKMN 860  
 QY 1027 GGLAAVSTSVNLSDQVLASRFMAVHKPLKNSODYTEALBARBELAANITADLRVPT 1086  
 DB 861 SPMLIYAKPFG-----ASRFMAVHTPLNSQETGALMKARBELAHNITAMARQIOGT 912  
 QY 1087 DBAFEVFPYT-----ITNVFEQYLTILPEGLF 1114  
 DB 913 DNFVEVFPYTYVLSFISLPLSHYSCDTFIHNDCVSPSLPARVTNVFYEGYVTVIPEGVL 972  
 QY 1115 MSLSLVPTTPANSCLLGLDLSGLNLISYVILVDTYGFMAIMDISIRNASVLIWVGA 1174  
 DB 973 IISLCLPTFPVCCLLIGDLRSGLNLITIIIMITVDYGVMTLWGIIDYNAVALINLVYA 1032

QY 1175 VQMSVEFVSHITRSFALSTKPTWLEBAKATISMGSAVNPAGVAMTNLPGIILGLAKQL 1234  
 DB 1033 VQISVEFVSHMTRSFALSIKPTHERAKAETQOMCSAVFAGVAMTNLPGIILGLAKQL 1092  
 QY 1235 IOIFFRNLNLITLGLHGLVFLPVILSYVG 1266  
 DB 1093 IOIFFRNLNLITLGLHGLVFLPVILSYFG 1124  
 RESULT 5  
 Q9ULG3\_CRIGR  
 ID Q9JUG3\_CRIGR PRELIMINARY; PRT; 1277 AA.  
 AC Q9JUG3;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DE Niemann-Pick type C1 protein.  
 GN Name=NPCL;  
 OS Cricetulus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Cricetinae; Cricetulus.  
 OX NCBI\_TaxId=10029;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=ovary;  
 RX MEDLINE=20127678; PubMed=10660558; DOI=10.1074/jbc.275.6.4013;  
 RA Cruz J.C., Sugli S., Yu C., Chang T.-Y.,  
 RT "Role of Niemann-Pick type C1 protein in intracellular trafficking of  
 RT low density lipoprotein-derived cholesterol.";  
 RL J. Biol. Chem. 275:4013-4021(2000).  
 DR EMBL; AF182744; AAF31692.1; -; mRNA.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0005764; C:lysosome; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008158; F:hepato receptor activity; IEA.  
 DR GO; GO:0030301; P:cholesterol transport; IEA.  
 DR InterPro; IPR004765; NP\_C type.  
 DR InterPro; IPR003392; Patched.  
 DR InterPro; IPR000731; SPD\_5TM.  
 DR Pfam; PF02460; Patched; I.  
 DR TrEMBL; TIGR00917; ZA060601; 1.  
 DR PROSITE; PS50156; SSD; 1.  
 SQ SEQUENCE 1277 AA; 142817 MW; 6B3D638094B4E156 CRC64;  
 Query Match 34.9%; Score 2413; DB 2; Length 1277;  
 Best Local Similarity 39.8%; Pred. No. 1e-161;  
 Matches 521; Conservative 245; Mismatches 438; Indels 104; Gaps 26;  
 QY 11 LVALLLRLAQSPEYTHIQPGYCAFYDEGKQNPILSGIMTLVNSVCLSNTPARKITGDH 70  
 DB 8 LGLILLILCPAIVFSQ-----SCWVYGC-----GIAFGKXKNCKSGSPKPLPKDG 55  
 QY 71 LILLOKICRPLTYGNTQAQCSAKQLVSLASLITKALLTRCPACSDNFVNLHCNHTCS 130  
 DB 56 NDLLQELCPGFPFG-NVSLICDVOQLQTLKSNLQIPMGELSRCPSCFYVLMILFCELTS 114  
 QY 131 PNOSLPIVNTR-----VAQLGAGQLPAVVAVEAFYGHSPAEQSYDSCSRPRVNAATLAVG 186  
 DB 115 PHOSQFLVNTATEDYVDPKTOENKTNVKELEYIQOSFANEMYNACRDVEAPASNEKALG 174  
 QY 187 TWCQVYGSALCNAORMLNFGDGTGNGLAPLDTIT-----FHLPEQAVSGGIQPLNEGVA 241  
 DB 175 IICGADASA-CAHTNWEIEMFKANNQAPFTITPISDLPII-----GNEPRNARK 225  
 QY 242 RCHESQGDVATCSQCDCAASC-----PAIARPOLDSTFYLGQMGPSLVLIILIC 292  
 DB 226 GCNSESVDVETGRCSCDCSIVCGPKPQPPPTVPWRILGLDAMY-----VIMWV 274  
 QY 293 SVPAVVTIILVFRVAPARDKSKMTPDKKGTSLSDKLSGHTT----- 335  
 DB 275 TYMAFLFIFFGGLLAVWCHRRRYFV--SEYTPIDSNIAFSSNTSDKGAASCCDPLGAAPD 332

QY 336 -LLGPFQMGKTWVAWMLTILVLSIYVVALAAGVTELTTPDPELMSAPNSQARSEK 394  
DB 333 DCLRMFTKMGAFVYRANPTCVIFSLVFTTACSSGGLVFRVTTNPVELMSAPHSRRAREK 392  
QY 395 AFHDQHFQFPRTNOVILTPAPNRSRYRPSDLLGPK-NFSGLDLDLELELBERLH 453  
DB 393 EYFDHFQFPFMEQLIAPNTSEKIYPPYPSGVSFGPPLNKEILHQVLDLQAIY- 451  
QY 454 LQWSPDARNTSLQDICYAPINPDNTSLYDCCINSLLQYFQNNRTLLLTANQTLMGOT 513  
DB 452 -SITTSYNNKVTLODICYAPLSPYK--NCTIISLVNYPONSHSVL-----DHQVGD 502  
QY 514 SQV--DMKDHELYCANAPLTFQDGTALASCMADYAPRPPLLAGYGYKQYSEAEAL 571  
DB 503 FYVYADYHTHFLYCVAPAPTSINDTSLDPCGTFEGGPVFWLVGGYDQYNNATYLV 562  
QY 572 MTFSLANNYPADGPRLAQAALMEAEFLBEMRAFORRAAGVQVTFAPERSLEDEIRTRAE 631  
DB 563 ITFPPSYNNDTEKIQRAQAMEKEFIDPKSYKN--PMLTISFLAERSIDELNRESN 619  
QY 632 DLPFATSYIYIPLYISLALGSYSWSRVVSDSKATLGLGVAVVLGVAAMAGFPSTYK 691  
DB 620 DVFTIAYSIAIMFLYISLALGHSKCSRLVDSKISLAGIILIVLSVACSLGVFSYWG 679  
QY 632 IRSSVLIVQVPELVLSVADNITFVLEYORLPRRGGPRRPHGRALGRAPSMILCS 751  
DB 660 MELTILIVLEVPFLVLAAGVNDIFILVQYQDERQESTLQOQGRILGEVAPFTMFLSS 739  
QY 752 LSEALCFGLATPMPAVRTFALTSGLAYILDLQMSAFVALLSDSKROBASRLDYVC 811  
DB 740 PSETAPFFPGALSSAPAVHTFSLFAGLANVLDLQITGFVSLGLGDIDRQKRNLDIIC 799  
QY 812 CYKPEQLPBPQGG---EGLLLGFPQKAYAPFLMHTTRGVLLFLALFGVLSYSMCHI 867  
DB 800 CVGGTD--NGRGIQASESYLFRFPKNSPAPFLKMDLRPIYIAVFGVLSPSIAMNKV 856  
QY 868 SVGLDQOELALPKDSYILDLFLNRYEFGAVYPTTIGVNFSEBAGNALLCSSAGCN 927  
DB 857 EIGLQOISMPPDSYVIDYFKSLGOLHSGPPVYFYLEBGRHYTTHKGQNMVCGMGCDN 916  
QY 928 FSEFQKIQYATEFPEQSYALAPASSVWDDFIMLTP--SSCCRLYISGPKDKFCPSTVNS 986  
DB 917 DSLVQOIFPAALDLNTRIGFAPSSVIDYFQWVAPQSSCCLLYNA--THQPCNASTVD 973  
QY 987 INCLAKNSIT-MGSVPSVEQFHKYLPWFLNDRPNIKCPKGLAAYSTVNLTSQG-QY 1044  
DB 974 PTCIR-CRPLTEGKORPQCKEFMKFLPMFLSDNPVKCKGKGHAAYSANVIGDDTV 1032  
QY 1045 LASRPMATHKPLKNSQDYTEALRAARELANITADLRKPGDPAEVEVPYITTVFQY 1104  
DB 1033 GATYPMYHTVTKTSADYIDAKKQQLVARNITETVNS-KGSN--YRVPYSPVYVFEQ 1089  
QY 1105 YTLIPBGLFMLSCLVPTFAVSCLLGLDLSGGLNLTSLVILVDYTGFMALMDISYN 1164  
DB 1090 YLTIIDDTFNSVLSGISFLVTVLVGLGELMSAVIMCTIMILVNMFGVWMLDISN 1149  
QY 1165 AVSLINLVSAVGSEVESHITRSFAISRTKPTWLEBAKATISKGSAPVAGVAMTNLPCI 1224  
DB 1150 AVSLVNLVMSGSIYVEFCSHITRAFTMSTKGSRVSRABALAHMGSVPSGITLTFGSI 1209  
QY 1225 LVYGLAKQQLIYFFRNLMLTLTGLGLGVFLPYILSYGPDVWPA 1272  
DB 1210 VYLAFAKQIFEIFYFRMYLAIVLGATHGLIFLPVLLSYGSPVKA 1257

RESULT 6  
NPC1\_PIG STANDARD; PRT; 1277 AA.  
AC P56941;  
DT 30-MAY-2000 (rel. 39, Created)  
DT 30-MAY-2000 (rel. 39, Last annotation update)  
DT 10-MAY-2005 (rel. 47, Last annotation update)  
DE Niemann-Pick C1 protein precursor.

GN Name=NPC1;  
OC Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
OX NCBI\_TaxId=9823;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21655028; PubMed=11796528; DOI=10.1210/en.143.2.708;  
RA Gevery N., Lacroix D., Song J.H., Pescador N., Doblas M., Murphy B.D.,  
RT "Porcine Niemann Pick-C1 protein is expressed in steroidogenic tissues  
RT and modulated by camp-";  
RL Endocrinology 143:708-716(2002).  
CC -!- FUNCTION: Involved in the intracellular trafficking of  
CC cholesterol. May play a role in vesicular trafficking in glia, a  
CC process that may be crucial for maintaining the structural and  
CC functional integrity of nerve terminals (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Found in late  
CC endosomes and lysosomes (By similarity).  
CC -!- DOMAIN: A cysteine-rich N-terminal domain and a C-terminal domain  
CC containing a di-leucine motif necessary for lysosomal targeting  
CC are critical for mobilization of cholesterol from lysosomes (By  
CC similarity).  
CC -!- SIMILARITY: Belongs to the patched family.  
CC -!- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
DR EMBL; AF169635; AAD47090.1; -; mRNA.  
DR InterPro; IPR004765; NP\_C\_Type.  
DR InterPro; IPR003392; Patched.  
DR InterPro; IPR000731; SSD 5TM.  
DR Pfam; PF02460; Patched; 1.  
DR TIGRfam; TIGR00917; 2A060601; 1.  
DR PROSITE; PS50156; SSD; 1.  
KW Glycoprotein; Lysosome; Signal; Transmembrane.  
FT SIGNAL 1 22  
FT CHAIN 23 1277 Potential.  
FT TRANSMEM 271 291 Niemann-Pick C1 protein.  
FT TRANSMEM 352 372 Potential.  
FT TRANSMEM 623 643 Potential.  
FT TRANSMEM 655 675 Potential.  
FT TRANSMEM 685 705 Potential.  
FT TRANSMEM 761 781 Potential.  
FT TRANSMEM 834 854 Potential.  
FT TRANSMEM 1039 1119 Potential.  
FT TRANSMEM 1125 1145 Potential.  
FT TRANSMEM 1196 1216 Potential.  
FT TRANSMEM 1228 1248 Potential.  
FT DOMAIN 620 785 SSD.  
FT MOTIF 1274 1277 Di-leucine motif.  
FT COMBINS 249 259 Poly-Pro.  
FT CARBOHYD 70 70 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 122 122 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 135 135 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 185 185 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 222 222 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 415 415 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 452 452 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 459 459 N-linked (GlcNAc...).(Potential).  
FT CARBOHYD 478 478 N-linked (GlcNAc...).(Potential).  
SQ SEQUENCE 1277 AA; 141963 MW; 2C80D300889F02EB CRC64;  
Query Match 34.9%; Score 2412; DB 1; Length 1277;  
Best Local Similarity 39.8%; Pred. No. 1.2e-161;  
Matches 524; Conservative 234; Mismatches 450; Indels 110; Gaps 25;  
QY 7 RGNLMLALLLALQSEPTTTHQPCYCAFYDECGNPELSGSLWTLNSVSCLSNTPARKI 66



Db 4 RGPAGLLLLLCPPVQFSQ-----SCWTGECG---IASGD-----KRYNCKYSGPKPL 51  
 QY 67 TGDHLLILKICIBRLYTGPNTQACCSAKQVLSEASLITKALITRECPACSDNFVNLHCH 126  
 Db 52 PEDGVYLVQELCGFFFG--NVSLCCDVQQLRTLKQNLQPLQFLSRCPSCFYVLMKLFCE 110  
 QY 127 NTCSPNQSILFINTR-----VAQIAGAGQLPAVVAEAFYQHSFPAEQSYDSCSRVRVPAAT 182  
 Db 111 LTCSPRQSQFLNLTATIEDVDPVTNQTKNVKELEFYVEGETFANAMYNACRDVEAPSSNE 170  
 QY 183 LAVGTGCGYYSALCNAQRMLNFQGDGTGKGLAPLDIT--FHLEPGQAVGSGIOPINEGV 240  
 Db 171 KALGLLGGREBAQH-CNATWIEIYFMKNQKQAPFTTPIFSDLPT-----HGMERNMNA 224  
 QY 241 ARCNSQGDVDVATCSQDCQACSC-----PALARPOLDSTFYLGQWPGSLVLIIL 291  
 Db 225 KCGDESVDEVYTGCSQDCSIVCGPKPPPPVPRILIGLDMYIMMSSWMAFLIVVF 284  
 QY 292 CSVFAVVTLLVGFRAVAPARQSKMYDPKKGTSLSQKLSFTHTL----- 336  
 Db 285 GAFFAVWCY-----RKRYFVSEYTPIDGNIAFVSNSDKQAFCCDPLGAAP 331  
 QY 337 ---LQFGQCGMTWVMSWPLTIVSVIPVMAAGLVPTLTTDVEILMSAPNSQARSE 393  
 Db 332 ERLRLFAQWGAFCVRHPCGVFFSLAFIVACSSGLVFRITTDVDDLMSAPSGARRB 391  
 QY 394 KAFHDQFGPPFPRTNOVILTPARSSRYRDSLLGPKR-NPSGILDDLLELLELLEORLR 452  
 Db 392 KEYFDTHFGPFRFMEQILIRATNOSHITHPRYAGADVPGPRLSDILHQVLDLQTAIE 451  
 QY 453 HLOQVSPERQNRISLQDICAPRLNPNTSLYDCCINSLLQYFQNNRTLLLTANQTLMOQ 512  
 Db 452 N--ITASYNNEVTTLQDICIAPLSPYNK--NCTILSVINFYQNSHVL----- 495  
 QY 513 TQGV-----DMKHFLYCANARPLTFDGTLLALSQADYCARPPPLIATGKYGKQY 564  
 Db 496 DHQVGFPEFVYADYHHPFLCYRAPASINDASLLHDPCLGTGFPVPMVLVCGYDDQNY 555  
 QY 565 SEAEALIMTFSLNNYPADGPRLQAQKLMEAFLEEMKRAFQRRMAAGFOVYTFERSLEBE 624  
 Db 556 NATALVITFPVNNYNDREKQRAQAMSEFIPNPKANYK--PULTISFMAERIEBE 612  
 QY 625 INRTTABDLPIFATSYIVFLYISLALGYSYSSWSRVVNDKATLGLGVAVALGAVMAAM 664  
 Db 613 LNRSNSDLFTTILISVIMFLYISALGHKSCSRLLVDSKISLGIAGILIVLSVASCIL 672  
 QY 685 GFPSYIGIRSSLVILQVNPFLVLSVQADNFIYVLEQYLRPRPGPREVHTIRALGRVA 744  
 Db 673 GIFSYIGVPLTLVIVIPFLVAVGVNDIFILVQTYQRDERLQGBETLDOQLGRVIGEVA 732  
 QY 745 PSMILCSLSEALCFEFLGALTMPMPAVRTPALTSGLAVILDFLQMSAFVALLSISDSKROA 804  
 Db 733 PSMFLSFSBETVAFIAGLSVVPRAVHTFSLFAGMAVILDLQITCFVSLISGLDIRQEK 792  
 QY 805 SRLDVCCKYPOELPPPGQG---EGLLIGFFQKAVAPFLMHTITRGVULLFLALFGVS 860  
 Db 793 NRLDVVCCVQGAH---DGAQVQASESCLPFFFGNSAPALLKQWMPRIYAVAVGLSS 849  
 QY 861 LYSMCHISVGLDOELAPKDSYLLDYFLFLNRYFENGAPRYFTTTLGYNFSSAGANNAIC 920  
 Db 850 IAVLNKYEIGLDQSLMPDQSYMDYFQSLSRYLHAGPPRYFVVEEGHNYTSLKQGNMVC 909  
 QY 921 SASACNNSFTQKIQVATERPEQSYLAPASSVNDPFIWLTP--SSCCRLYISGRPKDXF 979  
 Db 910 GGLGNCNDSLVOQIFTPAQDLNNTIRIGFAPSSWIDYFDMIKQSSCCRYSN---TDQF 966  
 QY 980 CBSYVSLNCLKNQMSITM-GSVRPSVEQFHXYLPFLNDRPNIKGPKGGGLAAYSTVNL 1038  
 Db 967 CNASVVDPCIR-CRPLTISEGKQRPQGEDPMRFLPMFLSDNPNPKGKGHAAVYSAGVNI 1025  
 QY 1039 TSSGQ--VLASRFPAHYKPLKNSQDYTBALPAARELANITADLRKYPGIDPAEVEVPYTI 1097  
 Db 1026 LGNGSGVGAITYFMTYHTVLOASADFLDAMQKARLIASNIT--RTWGLEASSYRVFPPYSV 1082

QY 1098 TNVFFBOYITLIPBGLFMLSCLVPTFAVSCILLGLDLSGLNLSIYMLVDYVEMA 1157  
 Db 1083 FYVFEQYVLTVIDDITFNLGSLGAI FLVTVLMGCELTATYIMCTIAMILLVNEGVMM 1142  
 QY 1158 LMDISYNAVSLINIVASVMSVEFVSHITRSPAFISKPMPLBRAKATISMGSAVPAVGA 1217  
 Db 1143 LMGISLNAVSLVNLVMSCGISVEFCSHITRAFTLSTKSGRVDAEALAHMSSVPSGILT 1202  
 QY 1218 MTNLPGLIVLGLAKQQLQIFPFFRLNLLITLLGLHGLVFLPVILSYGVPDVNPAAL 1275  
 Db 1203 LTKFGSIVVLAPAKSQIFQIFFFRMYLAVLLGATGGLFLFVLVLSIYSPINKASL 1260  
 RESULT 7  
 ID Q8MKD8\_FELCA PRELIMINARY; PRT; 1276 AA.  
 AC Q8MKD8\_ 08MKD8\_ 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 OC Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;  
 OC Felis.  
 OC NCBI\_TaxId=9685;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22695580; PubMed=12809639; DOI=10.1016/S1096-7192(03)00074-X;  
 RA Somers K.L., Royals M.A., Carstea E.D., Rafi M.A., Wenger D.A.,  
 RA Threlk M.A.,  
 RT "Mutation analysis of feline Niemann-Pick C1 disease";  
 RL Mol. Genet. Metab. 79:99-103 (2003).  
 DR EMBL: AF03634; AM27451.1; -, mRNA.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0005764; C: lysosome; IEA.  
 DR GO: GO:0016020; C: membrane; IEA.  
 DR GO: GO:0008158; F: cholesterog receptor activity; IEA.  
 DR GO: GO:0030301; P: cholesterol transport; IEA.  
 DR InterPro: IPR004765; NP C type.  
 DR InterPro: IPR003392; Patched.  
 DR InterPro: IPR000731; SLD\_5TM.  
 DR Pfam: PF02460; Patched; 1.  
 DR TIGRPFAM: TIGR00917; 2A060601; 1.  
 DR PROSITE: PS50156; SLD; 1.  
 SQ SEQUENCE 1276 AA; 141710 MM; C2DD3496472B71EF CRC64;  
 Query Match 34.8%; Score 2405; DB 2; Length 1276;  
 Best local Similarity 40.7%; Pred. No. 3, 8e-161;  
 Matches 535; Conservative 232; Mismatches 461; Indels 86; Gaps 29;  
 QY 7 RGLMLALLRLAQSPTTTHQPGCAFYDECGRKPELSGLMLTNSVCLSNTPARKI 66  
 Db 4 RGPAGLLLLLC---PAQVLAQS---CIWYGBCG---IASGD-----KRYNCKYSGPKPL 51  
 QY 67 TGDHLLILKICIBRLYTGPNTQACCSAKQVLSEASLITKALITRECPACSDNFVNLHCH 126  
 Db 52 PEDGVYLVQELCGFFFG--DNVSLCCDVQQLQTLKQNLQPLQFLSRCPSCFYVLMKLFCE 110  
 QY 127 NTCSPNQSILFINTR-----VAQIAGAGQLPAVVAEAFYQHSFPAEQSYDSCSRVRVPAAT 182  
 Db 111 LTCSPRQSQFLNLTATIEDVDPVTNQTKNVKELEQYIESFANAMYNACRDVEAPSSND 170  
 QY 183 LAVGTGCGYYSALCNAQRMLNFQGDGTGKGLAPLDIT--FHLEPGQAVGSGIOPINEGV 240  
 Db 171 KALGLLGGREBAQH-CNATWIEIYFMSKQNGAPFTTPIFSDLPT-----HGMERNMNA 224  
 QY 241 ARCNSQGDVDVATCSQDCQACSC-----PALARPOLDSTFYLGQWPGSLVLIIL 291  
 Db 225 KCGDESVDEVYTGCSQDCSIVCGPKPPPPVPRILIGLDMYIMMSTMAFLIVVF 284  
 QY 292 CSVFA-----VVTLLVGFRAV--PARDKSRMDPKKGTSLSDKLSFTHTLQO 339



Db	285	GAFPLMWCYRKXRYFYSEYTPIDSNIAFNVNADRGB-----ASCDAALGAAFBEGCLAR	337
OY	340	FFQMGWTVASWMPILITLIVLSVPEVVALAAGLYFTELTDPELMSAPNSQASSEKAFHQ	399
Db	338	LFSQMGSECVRNPGCIIFFPSLAFIAACSGLVFVRTPTVPDLMSPASQARLEKEFYDT	397
OY	400	HGPFEPFRINOVILTPNPNRSRYRSDLSLLGPK-NFSGDILDLLELLELOERLRHQWS	458
Db	398	HGPFEPFRREQULIQAPHISAHYTOPRPFSSDDVPFGPPLDALIHQVLDQTAIEN--TRA	455
OY	459	PEAGENISLQDICVAPLNEBDNTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQV--	516
Db	456	SYNNETVTLQDICVAPLSPYNK--NCTILSYLANYQNSHML-----DHEIGDPEFYVA	507
OY	517	DMKHFLYLCANAPLTFKQDTALALSGMDYGAVPVPELAGIGCYKGDYSEABRLNTBBL	576
Db	508	DYHTLLLCVVRPASILNDTSLIHLDPCLGTFGGVGFPEWMLVGGDDQNNATNALVITTFPV	567
OY	577	NNYPAGDRLIAQAKMEEAFLSEEMBAFQRMAGMFPVTFEASRLSEDEINRTAEDLPF	636
Db	568	NNYNDTERLQGAHWEKEFINFVKNYKN--PNLITSTTERSIDELNRENGDIFV	624
OY	637	ATSYVIVFLYISLALGSYSWSRVNVDSCATLGLGVAVALGAVMAAMGFPSYLGIRSSL	696
Db	625	IISYAIMPELYISIALGHIKSCRLLYVDSKISIGIAGILITLVLSVACSLGIFSYVGIPLTL	684
OY	697	VILQVVPFLVYSVGADNIFIPLEYQRLRRQEPREPHVIGALGVAAVSMLCSLEAI	756
Db	665	IYIEYIPELVAVGDNITLIVQTYQRDRLHGETIDQOGRVLGVAABSMFLSSSEAV	744
OY	757	CEFLGALTPMPAVRTFALTSGLAVIDLFLQMSAFVALLSLDSKQZASRLDYCCCVKQ	816
Db	745	APFLGALSMAPVHTFSLFAGMAVLIDPLTLQTCFVSLGLDIKRQEKRRLDVLCVRRGS	804
OY	817	ELPPPGQ--GEGILLGFQKAYAPFLMLHTTRGVULLPLALCGVSLYSXCHISVGDQEL	875
Db	805	EDGTSVQASESCLFRLFKHSYSPLLIKDMRRPLVIAI FGVGVSFSVAIVANKVEIGDQSL	864
OY	876	ALPKDSYLDYVLPFNRYFEVGAAPVYFVTTLGFNFSSEAGMNAICSSAGCNFSPQKIQ	935
Db	865	SMPDSDYVMDYKSL-KYLMHAPPVYFVLEBGHDYSLKGQNMVCCGMGNNDSLVQOJF	923
OY	936	YATEPPEOSYLAIPASSWVDFTDMLTP--SSCCRLYISCPNKDKFCPESTVNSLNCIANKM	994
Db	924	NAAQDSDYTRIGFAFSSWIDYDFDWKCPQSSCCRVNS--TDRFCNASVVPACIR--CR	979
OY	995	SITM--GSVRPSYEOGHXYLPWFLELNRPNKCKPKGGLAASTSVN--LTSQGVYLAASFMY	105
Db	980	PLTQSGKQPPQGGDFRFLPMFLSDPNKCGKGAAASAVANILGNTGCAITFKMY	103
OY	1053	HKPLKNSODYTEALRAARELIANITADLRKRVGTDPA--FEVPEYTTTNYFEYOYLITPE	111
Db	1040	HTVLQTSADFTDAMERKANLIASNIT---KTMGLEGSNRYVPFYSVYFVEQYGLITIDD	109
OY	1112	GLFMLSCLVPEFPFANVSCLLGLDLSGGLNLSIWMIVLDYVGFMAIMDISYAAVSLINL	117
Db	1096	TIFNLSVSLGALFELVTVILLGGDLSAVIMCTITAIILVNMFGVWMLMGSLNAAVSLNL	115
OY	1172	VSAVGMSVEFVSHITRSFAISTKPTWLBRKAKETIIMSAGAVFAGVAMTNLPGLIVGLAK	123
Db	1156	VMSGCISIEFCHITRAFVMSKKGSAQAQAEHLAMGSSVFSGITLTLRFGGIVLVAFPK	121
OY	1232	AQLOIIPFRRLNLLITLGLHLGLVFLPVILSYGVDPVNPALAL--EQKRAEE	1282
Db	1216	SOIPIPIFFRRMYLANVLGATGTLFLPVLILSTIGSINKAKSLAQEOBYKGE	1269

RESULT 8	
09GK52 CANPA	
ID 09GK52 CANPA PRELIMINARY;	PRT, 1276 AA
AC 09GK52-	
DT 01-MAR-2001 (TREMBlrel. 16, Created)	

DT 01-MAR-2001 (TREMBLrel. 16 Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Niemann-Pick type C1 disease protein.  
GN Name=NPCL1;  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
OC Canis.  
OX NCBI\_TaxId=9615;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Murakawa M., Freeman M.W.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF151034; AAG40873.1; -, mRNA.  
DR Ensembl; ENSCAF000000018183; Canis familiaris.  
DR GO; GO:0016021; C:Integral to membrane; IEA.  
DR GO; GO:0005764; C:Lysosome; IEA.  
DR GO; GO:0016020; C:Membrane; IEA.  
DR GO; GO:0008158; F:hedgehog receptor activity; IEA.  
DR GO; GO:0030301; P:cholesterol transport; IEA.  
DR InterPro; IPR004765; NP\_C type.  
DR InterPro; IPR003392; Patched.  
DR InterPro; IPR000731; SSD 5TM.  
DR Pfam; PF02460; Patched; I.  
DR TIGRfams; TIGR00917; 2A060601, 1.  
DR PROSITE; PSS0156; SSD; 1.  
SQ SEQUENCE 1276 AA; 141695 MW; 07D1327B51BD5681 CRC64;

Query Match	Similarity	34.8%	Score 2404.5	DB 2	Length 1276
Beet Local	Similarity	40.4%	Pred. No. 4.2e-161		
Match 524	Conservative	233	Mismatches 433	Indels 105	Gaps 28
QY	33	CAFYDECCKNELSGSLMTLSNVSCLSNTPAKRTGDIHLILLOKICPRLTYGPNTOACCS	92		
DB	25	CWVYGECC---IASGD---KRYNQYSGPKPEPKDQYDLMOGELCGIIEF-DNYSVCCD	76		
QY	93	AKQVLSLEASISITALLTRCPACSDNFWNLHCHNTCSPNOSLFINVR-----VAQLAG	148		
DB	77	VQQLRTLDSDSQPLQPLSRCPSCCYNLMNIFCELTCSPROSQFLNATVETEDYVDPVNTQ	136		
QY	149	QLPAVAVAEAFYOHSEFAEQYSDSCSRVAVPAAATLAVTMCVGSALCANQRMILFOGD	208		
DB	137	TKTNVKELQYVYVGSFANAMYNACRDVABPSNDALGLGCKAEV-CNANINLEMYEN	195		
QY	209	TGNGIAPL-----DITFHLLPEQAVSGEIQPLINEGVARCNEOGDDVATCSQODCA	261		
DB	196	KDNGGAPETIIPFIISDLPRAH-----GMKPMNNATKGCDEBEVDYETAICSQDCSCV	245		
QY	262	SC-----PALARPALDSTFYLGQMPGSLVLIILCSGYFAVTTLLVGFRAVPARD	312		
DB	246	VCGPKPQPPAPAPWRILGLDMAYIVIMITYMAFLMFFGAFAVWCY-----	293		
QY	313	KSXWDPKKGTSLSDPKLSFSTHT-----LLGQFFQG-----WGTWVASWPLT	354		
DB	294	RKRYFVSEYPIPIIDENIAFSVNAQDTGEASCCDNLGAAPFEGCLRRLFTQWGSFCIRNPQC	352		
QY	355	ILVLSVIEVVALAAGLVTELTDPVVELMSAPNSQASBEKAFHDOHFGPFFFTNOYLTLA	414		
DB	353	IIFELAIACSSSELVRSRYTINVDLMSAPGQARLEKEFEDAHFGFFFTEDLIIOA	412		
QY	415	PNRSYSYRDSLTLGCK-NFSGILDDLLLELELEOERLHLQVMSPEAQRNLSODICVA	473		
DB	413	PHTSVHTYQPIPSGSDVDFGPPLEDIGILHQYLDLQTALEN--ITATYNNETVTLDDICVA	470		
QY	474	PLNPNTSLYDCCINSLLQYFQNNETLLLTANQTLMGQTSQV--DWKDEHLYCANAPLT	531		
DB	471	PLSPYNNK---NCTIMSVANYFONSHSMU-----DHKIGDDFYVADYHTHLLLYCVRAPAS	522		
QY	532	KFDGALMLSCMADGAVPEPRLAIGKXGQVSAEALLIMFSSNINPAGDPRLAQAAL	591		
DB	533	LNDTSLHDPCLGTFGGFVPEPLVAGGVDQONATNATLVITFVANNYINDTEKIORAQA	582		
QY	592	WEBALEEMRAFORRMAGMFOVTFPAERSLEDEINRTTAEDLPFATSYVIFLYISIAL	651		

Db 583 WEKEFIINFKVKNYEN---PNTLISFTTERSTIEDEINRESNDGVFVLIVSYAMVFLYSIAL 639  
 Qy 652 GSYSSNRVWVDSKATLGLGVAIVV.GAVMAAGFEFYLIRSRSLVLOVPPFLVYSVGA 711  
 Db 640 GHKISCRFLVDSKISLGIAGILVLIVSSVNCSCIFEFIPGLVLIVETVLPVLAVGV 699  
 Qy 712 DNIFIPLYLEYQRLPRRGEEREVHIGRALGRVAPSMILCSLSEATCFGLATPMPAVRT 771  
 Db 700 DNIFILVQYQROBERLOGFTLBQOLGVLEBVAWSMFLSFSEAVAFGLALSGMPRVHT 759  
 Qy 772 FALTSGIATLDPLOMSAFVALISLDSKQOASRLDYCCCVXPKQELPPRQG---EGL 827  
 Db 760 FSLPAGAAVVIDELQITCFVSLIGDLTKQEKRRDLVLCCLTGE---GATGLOASESC 816  
 Qy 828 LDGFPQAVAPFLHMTTRGVLLFLALFGVSLYSCHISVGLDQSLAPKXSYLLDYF 887  
 Db 817 LFRFFKNSVSPFLKDMRPVIAVFGILSFSTIAVINKVETIGDQSLMPDDSYMMDYF 876  
 Qy 888 LFLARYFEVGAPEVYFVTLTGYNFSSSEAGNNALCSSAGCNNSFTOKIQVATEPEQSYLA 947  
 Db 877 KSL-KYIHAGPRYFVLEBGHDYTSLEGOMVCGMKNNDLSVQOIFSAAGLDNTHIG 935  
 Qy 948 IPASSWVDFIDWLTTP-SSCCRLYISGPNKDCPCPTVNSLNCIKNCMSITM-GSVAPSV 1005  
 Db 936 FAPSSWIDYFDWVKPQSSCCRVYNS---TDQFCNAGSVDPACVR-CRPLTQSGKRPPQ 991  
 Qy 1006 EOHKXILPMFLNDRPNKCKPKGGLAASVTSVLT-SNGOVLASRFMAVHPRKNSQDYTE 1064  
 Db 992 EDPMRFLPMFLSDPNPKCKGGLAAGSAVNLGNSTSGATYFMVHTVLTQSAFETD 1051  
 Qy 1065 ALBAARELANITADLRKVPETDPAPAFEPYITTVNEVEQYLTLPFGLFMLSCLVPTF 1124  
 Db 1052 AMKATILIANITKTM-SLKGSH--YRFPYSVYVEYEQYLTITIDTINLSVSGAIF 1108  
 Qy 1125 AVSCLLGLDLRSGLLNLISVILVDTVGFMAIMDISYNAVSLINISVAGMSVEFVSH 1184  
 Db 1109 LVTVLVLGCELMSAVIMCWTIAMILVMFGVMMLMGISLNAVSLNLMSCGISVERCSH 1168  
 Qy 1185 ITTSFALSTPMLERAKENTISGSAVFPAGVAMTNLPGILVGLAAQLOIOPFFPLNL 1244  
 Db 1169 ITRAFVYSAGSRVERAEALSHNGSSVFSGITLTKEGVLVAFAKSQIFQVFTGMYL 1228  
 Qy 1245 LITLLGLHGLVFLPVLTSTVGPDPVNPALAL-EOKR 1279  
 Db 1229 AMVLGKTHGLIFLPVLTSTIGSINKAKSLAQOER 1264  
 RESULT 9  
 NPCL\_HUMAN  
 ID NPCL\_HUMAN STANDARD; PRT; 1278 AA.  
 AC 015118; 09P130;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 10-MAY-2005 (Rel. 47, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Niemann-Pick C1 protein precursor.  
 GN Name=NPCL;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 NCBI\_TaxID=9606;  
 RX MEDLINE=9362323; PubMed=9211849; DOI=10.1126/science.277.5323.228;  
 RA Carstea E.D., Morris J.A., Coleman K.G., Loftus S.K., Zhang D.,  
 RA Cummings C., Gu J., Rosenfeld M.A., Pavan W.J., Krizman D.B.,  
 RA Nagle J., Polymeropoulos M.H., Sturley S.L., Ioannou Y.A.,  
 RA Higgins M.E., Comly M., Cooney A., Brown A., Kanaseki C.R.,  
 RA Blanchette-Wackie E.J., Dwyer N.K., Neufeld E.B., Chang T.-Y.,  
 RA Liscum L., Straus J.F. III, Ohno K., Zeigler M., Carmi R., Sokol J.,  
 RA Markie D., O'Neill R.R., van Diggelen O.P., Ellender M.,  
 RA Patterson M.C., Brady R.O., Vanier M.T., Pentchev P.G., Tagle D.A.;

RT "Niemann-Pick C1 disease gene: homology to mediators of cholesterol  
 RT homeostasis.";  
 RT Science 277:228-231(1997).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE, AND VARIANTS.  
 RX MEDLINE=9355599; PubMed=10425213; DOI=10.1006/dbrc.1999.1070;  
 RA Morris J.A., Zhang D., Coleman K.G., Nagle J., Pentchev P.G.,  
 RA Carstea E.D.;  
 RT "The genomic organization and polymorphism analysis of the human  
 RT Niemann-Pick C1 gene.";  
 RL Biochem. Biophys. Res. Commun. 261:493-498(1999).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE, AND VARIANT ILE-642.  
 RX MEDLINE=21623216; PubMed=11754101; DOI=10.1002/humu.10016;  
 RA Bauer P., Knoblich J.P., Bauer C., Finckh U., Hufen A., Kropp J.,  
 RA Braun S., Kustermann-Kuhn B., Schmidt D., Harzer K., Rolfs A.;  
 RT "NPC1: complete genomic sequence, mutation analysis, and  
 RT characterization of haplotypes.";  
 RL Hum. Mutat. 19:30-38(2002).  
 RN [4]  
 RP TISSUE-PLACENTA;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,  
 RA Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ueban T.B., Toshiyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulyady S.J.,  
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smaluk D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Merra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=99128318; PubMed=9927649; DOI=10.1073/pnas.96.3.805;  
 RA Watari H., Blanchette-Wackie E.J., Dwyer N.K., Glick J.M., Patel S.,  
 RA Neufeld E.B., Brady R.O., Pentchev P.G., Straus J.F. III;  
 RT "Niemann-Pick C1 protein: obligatory roles for N-terminal domains and  
 RT lysosomal targeting in cholesterol mobilization.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:805-810(1999).  
 RN [6]  
 RP VARIANT NPD TRP-992.  
 RX MEDLINE=98299797; PubMed=9634529;  
 RA Greer W.L., Riddell D.C., Gillan T.L., Girouard G.S., Sparrow S.M.,  
 RA Byers D.M., Dobson M.J., Neumann P.E.;  
 RT "The Nova Scotia (type D) form of Niemann-Pick disease is caused by a  
 RT G3097-->T transversion in NPC1.";  
 RL Am. J. Hum. Genet. 63:152-54(1998).  
 RN [7]  
 RP VARIANTS NPCL GLN-934; LEU-940; ASN-948; LEU-954; TRP-992; ALA-1007;  
 RX THR-1061 AND VAL-1213.  
 RP MEDLINE=99452586; PubMed=10521290;  
 RA Greer W.L., Dobson M.J., Girouard G.S., Byers D.M., Riddell D.C.,  
 RA Neumann P.E.;  
 RT "Mutations in NPC1 highlight a conserved NPC1-specific cysteine-rich  
 RT domain.";  
 RL Am. J. Hum. Genet. 65:1252-1260(1999).  
 RN [8]  
 RP VARIANT NPCL THR-1061.  
 RX MEDLINE=99452593; PubMed=10521297;  
 RA Millet G., Marcalo C., Rafi M.A., Yamamoto T., Morris J.A.,  
 RA Pentchev P.G., Ohno K., Wenger D.A., Vanier M.T.;

RT "Niemann-Pick C1 disease: the 11061T substitution is a frequent mutant  
RT allele in patients of western European descent and correlates with a  
RT classic juvenile phenotype.";  
RL Am. J. Hum. Genet. 65:1321-1329(1999).  
RN [9]  
RP VARIANTS NPC1, AND VARIANTS ARG-215, VAL-858 AND GLN-1266  
RX MEDLINE=9408226; PubMed=10480349; DOI=10.1007/s004390051057;  
RA Yamamoto T., Namba E., Nishimura H., Higaki K., Taniguchi M., Zhang H.,  
RA Akapochi S., Watanabe Y., Takeshima T., Inui K., Okada S., Tanaka A.,  
RA Sakuragawa N., Millat G., Vanier M.T., Morris J.A., Pentchev P.G.,  
RA Ohno K.;  
RT "NPC1 gene mutations in Japanese patients with Niemann-Pick disease  
RT type C.";  
RL Hum. Genet. 105:10-16(1999).  
RN [10]  
RP VARIANTS NPC1 GLN-958 AND ALA-1007.  
RX MEDLINE=2131311; PubMed=11349211;  
RA Sun X., Marks D.L., Park W.D., Wheatley C.L., Puri V., O'Brien J.F.,  
RA Krafat D.L., Lundquist P.A., Patterson M.C., Pagano R.E., Snow K.,  
RT "Niemann-Pick C variant detection by altered sphingolipid trafficking  
RT and correlation with mutations within a specific domain of NPC1.";  
RL Am. J. Hum. Genet. 68:1361-1372(2001).  
RN [11]  
RP VARIANTS NPC1 ALA-378, MET-950, ARG-992 AND THR-1061.  
RX MEDLINE=21313105; PubMed=11333361;  
RA Millat G., Marcalo C., Tomasetto C., Chikh K., Fensom A.H., Harzer K.,  
RA Wenger D.A., Ohno K., Vanier M.T.;  
RT "Niemann-Pick C1 disease: correlations between NPC1 mutations, levels  
RT of NPC1 protein, and phenotypes emphasize the functional significance  
RT of the putative sterol-sensing domain and of the cysteine-rich luminal  
RT loop.";  
RL Am. J. Hum. Genet. 68:1373-1385(2001).  
RN [12]  
RP VARIANTS NPC1 TYR-177, CYS-978 AND VAL-1035.  
RX MEDLINE=21372069; PubMed=11479732; DOI=10.1007/s004390100531;  
RA Ribeiro I., Marcao A., Amaral O., Sa Miranda M.C., Vanier M.T.,  
RA Millat G.;  
RT "Niemann-Pick type C disease: NPC1 mutations associated with severe  
RT and mild cellular cholesterol trafficking alterations";  
RL Hum. Genet. 109:24-32(2001).  
CC -1- FUNCTION: Involved in the intracellular trafficking of  
CC cholesterol. May play a role in vesicular trafficking in glia, a  
CC process that may be crucial for maintaining the structural and  
CC functional integrity of nerve terminals.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Found in late  
CC endosomes and lysosomes.  
CC -1- DOMAIN: A cysteine-rich N-terminal domain and a C-terminal domain  
CC containing a di-leucine motif necessary for lysosomal targeting  
CC are critical for mobilization of cholesterol from lysosomes.  
CC -1- PTM: Glycosylated.  
CC -1- DISEASE: Defects in NPC1 are the cause of Niemann-Pick disease  
CC type C1 (NPC1) [MIM:257220]. NPC1 is an autosomal recessive lipid  
CC storage disorder, which affects particularly the brain, liver and  
CC spleen, and which is characterized by lysosomal accumulation of  
CC low density lipoprotein derived cholesterol. Clinical features  
CC include variable hepatosplenomegaly and severe progressive  
CC neurological dysfunction such as ataxia, dystonia and dementia.  
CC The age of onset can vary from infancy to late adulthood.  
CC -1- DISEASE: Defects in NPC1 are the cause of Niemann-Pick disease  
CC type D (NPD) [MIM:257220], also known as Niemann-Pick disease  
CC without sphingomyelinase deficiency, or Nova Scotian type. Because  
CC of evidence from biochemical changes, lack of complementation, and  
CC linkage mapping to the same chromosome site, NPD and NPC1 are  
CC considered to be allelic disorders.  
CC -1- SIMILARITY: Belongs to the patched family.  
CC -1- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.

DR EMBL; AF002020; AAB63982.1; -; mRNA.  
DR EMBL; AF157379; AAD48006.1; -; Genomic DNA.  
DR EMBL; AF157365; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157366; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157367; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157368; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157369; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157370; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157371; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157372; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157373; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157374; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157375; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157376; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157377; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157378; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF338230; AAK25791.1; -; Genomic DNA.  
DR EMBL; AF123045; AAF28875.1; -; Genomic DNA.  
DR EMBL; AF123045; AAF28875.1; JOINED; Genomic DNA.  
DR EMBL; BC063302; AAF63302.1; -; mRNA.  
DR Ensembl; ENSG00000141458; Homo sapiens.  
DR HGNC; HGNC:7897; NPC1.  
DR MIM; 607623; -.  
DR MIM; 257220; -.  
DR GO; GO:0016021; C: integral to membrane; TAS.  
DR GO; GO:0005764; C: lysosome; TAS.  
DR GO; GO:0005624; C: membrane fraction; TAS.  
DR GO; GO:0005478; F: intracellular transporter activity; TAS.  
DR GO; GO:0015248; F: sterol transporter activity; TAS.  
DR GO; GO:0004888; F: transmembrane receptor activity; TAS.  
DR InterPro; IPR004765; NP\_C\_Type.

Query Match 34.8%; Score 2402.5; DB 1; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 5.8e-161;  
Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

7 RGMALLLRRLRAGEPPTTHIQPGYCAFYDECG-----KNPLSGSLMTLSVNSCLS 59  
4 RRLALGGLLLLLCPAIVFSQ-----SCWYGGCGIAYDKRYNCYSG----- 46  
60 NTPARKITGDHLLLOKICPRLYTGPNTQACSAQOLVLSIASITLALTRCPAGSDN 119  
47 --PKKPLPDGDLVQELCPGFPG--NWSLCCDVRLQTLKNDLPLQFLSRCSCTYN 103  
120 FVNLCHNTCSBNGLFNTVR---VAQLAGQQLPAVAVYEAFFYQHSFAEGSYDCSRV 175  
104 LNLRCFELTCSRQGLFNTVEDYDPTVNTQTKNVEQLQYGGSFANMAYNACRDV 163  
176 RVPAAATLAVGCMGVYSALCNAGRWLNFQGDITNGSLAPLDIT-----PHLEPQQA 228  
164 EAPSSNDKALGLLCKGKDA--CNATNWIEMFNKNGGAFTITVFSDFPVH----- 215  
229 VQSGIOLPLNEGVARCNESGQDDVATCSQDCAASCPALARPQ-----ALDSTFYLG 279  
216 ---GHEPPNNAITKGCDESDVETALPCSCODCSIVCGPRPQPPRAPPTITGLDMYIM 272  
280 QMPGSLVLIILCSVPAVVTILL-----VGFVAPARDKRMVDPKKGTSL 326  
273 WITVAFLVLFVGAFPAVWCRAKRYFSEYPTIDSNIAFSV--NADKGS-----ASCC 324  
327 DKLSTSTHTLLGQFPQSGMTWVAPLTIIVLSVPVVALAAGLVETLTDPVELMAGP 386  
325 DVSAVAFECRCRLRTRGSPFCVRNRPGVIFPSLVFTRACSSGLVFVVTNNPVDLWAP 384  
387 NSQASSEKAFHGHGPPRTNQVLTLPNNSVSYGSLLLGPK--NFSGLIDLDLLELL 445  
385 SSQALREKYEFGHQHGFPRFTQLITRALDVKIYQPPSGADVPFGPEPDIDQLHQL 444  
446 ELQERLRLHQVSPBAONISLQDICVAPLNDNTSLVDCINSLLQYFQNNRTLLLLTA 505  
445 DLQALIN--IYASDNTETVLTQDICLAPLSVYNT--NCTILSVLNTFQNSHSLVDHKK 499  
506 NOTLMGQTSQVMDKPHLYCANAPLTFQDGTALALSCADYGAVPFPPLAIGYKGYDS 565

Db 500 GDDFF--VYADYHHTFLYCVRAPASLNDTSLHDPCLGTFGGPVPFWLGGYDDQNN 556  
 566 EBAALIMTFSILANNYPAGDPRLAQAOKLMEBAFLBEMKAFORBMAGMGOVTFTARSLEDEI 625  
 557 NARLALVITFPVNNYNDTEKQRAQAMEKEFIVKYNK---PNLTISTARSJEDL 613  
 626 NRTTADLPFATSYVIFLYISLALGSYSWSRWVDSKATLGLGVAVLCAVMAAMG 665  
 614 NRESDDVDFVTVAISYALIMFLYISLALGHMKSCHRLVDSKVSIGIILIVLSSVASCIG 673  
 686 FRESYLGRSLVILQVAVPELVLSVGADNIFILEYORLPRRGEPREHVIIGALRVAP 745  
 674 VFSYIGLPLTVILEVIFPLVAVGVNIFILVQVQDRERLQGETLDQQLGKVLGBVAP 733  
 746 SMLCSLSEALICFGLALTPMPAVRTPALTSGLAVILDLQMSAFVALSLSKQOAS 805  
 734 SMLSSFSSEVAFPLGLALSMVPAVHTPSLFCAGLAVFIDFLQITCFVSLGLDIIKQOEN 793  
 806 RLDDVCCCVKQBELPPPGQ-GEGLLGFQKAYAPFLHMTTRGVLLFLALRGVSLYSM 864  
 794 RLDIFFCVGAEDGTSVQASESCLEFRFFKNSYSPLLKDMWRPIVIAIFGVLSFSIAVL 853  
 865 CHLSVGLDQELAPKDSYLLDYFLPLNRFEVGAPEYFTVTLTGYNSSSEAGMNAICSSAG 924  
 854 NKVDIGIDQSLSMRSDSYWVDYFKSISQYLHAGPPVYFLEBGHDYSSKQNMVCGMG 913  
 925 CNNSFPTOKIQAATEPREGSYLAIPASSWVDDFIDMLTP--SSCCRLYISGNPKDKCPST 983  
 914 CNNDLSLVQQLFNAAQOLDNNTYRIGRAPSSWIDYDFDMKPOSSCCRV--DNITDQFENAS 970  
 984 VNSLNCNKMSIT--MGSVAPSVQEFHKLPMFLNDRPNIKCPKGLAAVSTSVN--LTS 1040  
 971 VVDPACR-CRPLTPBCKQRQGGDFMRFLPMFLSDNPNKCGKGGHAAVSSAVNLLGH 1029  
 1041 DGOVLASRFAAYHKPLKNSODYTEALPAEELANITADLRKYPGIDPAEVPYITTV 1100  
 1030 GTFVGATYFMTYHTVLTQTSADFDLAKKARLIASNTY-ETMGINGS--AYRVPYSVFYV 1086  
 1101 FYBOYLTILPEGLFMLSCLVPPFAVSCLLGLDLRSGLLNLSTIWIIVDTVGPALMD 1160  
 1087 FYBOYLTITIDTIFNLGVSIGALFVLVWVLGCELSAVIMCATIAWLVNMGVMMKMG 1146  
 1161 ISYNAVSLINLVAAGSVSEFVSHITRSPAISTKPTLBERAKBATISMGSAVAGVAMTN 1220  
 1147 ISLNAVSLVNLVWMSGIVSEFCSHITRAFVYSMKGSVERABEELAHMSSVSIGITLK 1206  
 1221 LPEILVIGLAKAQLIQIFFFRLLMLITLGLLHGLVFLPYITLSYGPVNPA 1272  
 1207 FGGIVVLAPAKSOIFQIFFRMYLWMLVLTGATGLFLPVILSYIGPSVAKA 1258  
 RESULT 10  
 059GRL HUMAN PRELIMINARY; PRT; 1289 AA.  
 AC 059GRL;  
 DT 10-MAY-2005 (Tremblrel. 30, Created)  
 DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)  
 DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)  
 DE Mlemin-pick disease, type C1 variant (fragment).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Brain;  
 RA Totori Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,  
 RA Ohara O., Nagase T., Kikuno F.R.;  
 RT "None Title."  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB209048; BAD92285.1; -; -;

FT NON\_TER 1 1  
 SQ SEQUENCE 1289 AA; 143230 MW; 663DE27C13FD801F CRC64;  
 Query Match 34.8%; Score 2402; DB 2; Length 1289;  
 Best Local Similarity 39.7%; Pred. No. 6,3e-161;  
 Matches 525; Conservative 229; Mismatches 452; Indels 106; Gaps 26;  
 2 AENGLRGWLLMA-----LLRLAQSPEYTTIHQRCYAFYDECG-----KPELSG 47  
 3 ASAAALRGHSMTRAGLALGLLILLCPAQVFSQ-----SCVMYGECCIAAYGDKRYNCEYSG 57  
 48 SLMTLSNVGCLNTPRARKITGDLILLLOKICRCLYTPPTQACCSAKOLVSLASLSIRK 107  
 58 -----PEKPLPKDGYNDVQBLCEGFFEG-NVSLCCDVRLQTLKDNLQLEL 102  
 108 ALITRCPSADNFVNIHCNTCSPNQSLFINYTR-----VAOLGAGLPAVVAVEAFYQHS 163  
 103 QPLSRCPSCFYLLNLNFCBLTCSPPQSPPLANTABEDYVDPTYNQKTNVKELQYVVGGS 162  
 164 FAEQSYDSCSYRVVPAAATLAVGTMCGVYGSAICNAORMLNFQDGTGNGLAPLDT---- 219  
 163 FANAMYNACRDVEAPESNDKALGLCGKQDAD-CNATNMIERYMKNQGAQAFITIPVPS 221  
 220 -FHLEPGQAVSGIOLPMBGVARCNESGDDVATCCODCAACPAIARPO----- 270  
 222 DPPV-----KGMENNNATKGCDESYDEVTAIPCSCDSIVCGPKQPPPPAPWTI 273  
 271 -ALDSTFYLGOMPGSLVLIILISVPAVVTILL-----VGRVAPARDKSKM 316  
 274 LGIDANVTYIMVITTYNAFLVFGGAFPAVWCYKRRYVSYTYTIDSIASV-NASDKG- 331  
 332 -----ASCCDPVSAAFEGCLRLRLFTRWGSCFVRNPGCVIIPSLVITRACSSGLVEVRVY 385  
 317 VDPKKTSLSDKLSFTHTLLGQFPQMGTVWASWPLTILVLSVIPVVALAAGLVETELT 376  
 377 TDPVELMSAPNGQASERKAFHQQHGGPFRFTQVILITANRSSYKRDLSLLPK-NFSGI 435  
 386 TNPVDLMSAPSSQARLEKEYFDQHPGPFRTQLIIRALTDKHIQYPSGADVEFGPP 445  
 436 LDDLLLELLELOERRLHQVMSPEAQRNLSIODICYAPLNPNTSLYDCCINSLHQYQ 495  
 446 LDIQLIHQVLDLQIAIEN--ITASYNETVITODICLAPLSPTN---NCTILSVANTYQ 500  
 496 NNRLLLLTANQTLQMTQSDVMKDFLYCANAPLTFKDGITALALSQMDYGAPEPFLA 555  
 501 NSHSVLDHKKGDFF--VYADYHHTFLYCVRAPASLNDTSLHDPCLGTFGGPVPFWL 557  
 556 IGGYKGDYSEBALIMTFSILANNYPAGDPRLAQAOKLMEBAFLBEMKAFORBMAGMGOVT 615  
 558 LGGYDDQNNYNNATVITPVPNNYNDTEKQRAQAMEKEFINFVKNYK---PNLTISF 614  
 616 TARSLEDIRNRTAEDLPFATSYVIFLYISLALGSYSWSRWVDSKATLGLGVAV 675  
 615 TARSLEDIRNRESDDVFTVVAISYALIMFLYISLALGHMKSCHRLVDSKVSIGIIL 674  
 676 VLGAVMAAMGFSYLGIRSLVILQVAVPELVLSVGADNIFILEYORLPRRGEPREHVA 735  
 675 VLSVASCISGVSYIGLPLTVILEVIFPLVAVGVNIFILVQVQDRERLQGETLDQQLG 734  
 736 IGRALGRVAPSMILGSLSAICFGLALTPMPAVRTPALTSGLAVILDLQMSAFVAL 795  
 735 LGRVIGEVAPSMFLSSFSSEVAFPLGLALSMVPAVHTPSLFCAGLAVFIDFLQITCFVSL 794  
 796 SLDSKQOASRLDDVCCCVKQBELPPPGQ-GEGLLGFQKAYAPFLHMTTRGVLLFL 854  
 795 GLDIKROEKNRLDIFCCVGAEDGTSVQASESCLEFRFFKNSYSPLLKDMWRPIVIAIFV 854  
 855 ALFGVSLVSMCHISVGLDQELAPKDSYLLDYFLPLNRFEVGAPEYFTVTLTGYNSSSA 914  
 855 GVLSFSIANLVNKVDIGDQSLSMRSDSYWVDYFKSISQYLHAGPPVYFLEBGHDYTSK 914  
 915 GMAATSSAGCNNSFTOKIQAATEPREGSYLAIPASSWVDDFIDMLTP--SSCCRLYISG 973

DB 915 GQMMVCGMGCGNNDLVQOIFNNAQUDNTRIGFASSWIDYFDPVYKQSSCCRV--D 971  
QY 974 PKDKRCPSVNSLNLCKNCKMSIT-NGSVRPSVEQFHXYLPFLNDPRNPKCKGGLAAY 1032  
DB 972 NITDGCNANSVDPACVR-CRPLTPRGKQRPGCGDMRFLPMLPSNPNPKCKGSHAY 1030  
QY 1033 STSVN--LTSQGVYLAASRPMAYHKPLKNSODYTEALRAARELANITADLRKYPGDPAF 1090  
DB 1031 SSAVNTLHCHGTRGVATYFMVYTYVTVQTSDAFDAKKARLIASNT-ETMGINGS--AY 1087  
QY 1091 EYEPYITNVFEYQVYITLIPBGLFMLSCLVFPFAVSCILGLDASGLNLTSTMYIV 1150  
DB 1088 RVFPFSVFEYFEQYITLIDTITLNLGSLGALFVTVWVLGCELSAVIMCATTMVAV 1147  
QY 1151 DTVGFMAIDISYNAVSLINLVSAVGSYEFVSHITRSPALSTKPTWLERAKEATISMG 1210  
DB 1148 NMFGVWMLGSLNANSLVNLVMSGSIYSVEFCHITRAFTVSKSQRVERABEALAHMS 1207  
QY 1211 ANPAGVAMTNLPGLVLGLAKQQLIIFFRNLTLITLGLHGLVFLPVILSYGPDV 1270  
DB 1208 SVFSGITLTKFGIVVLAFAKSOIFQIFYFRMYLAVMLGATHGILFELVLLSYIGPSV 1267  
QY 1271 PA 1272  
DB 1268 KA 1269  
RESULT 11.  
Q7TMD4 MOUSE  
ID Q7TMD4\_MOUSE PRELIMINARY; PRT: 1277 AA.  
AC Q7TMD4;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, last annotation update)  
DE Niemann Pick type C1.  
DE Name=Npc1;  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stachewicz M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodighiero S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Maita M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;

RA Strausberg R.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC052437; AAH52437.1; -, mRNA.  
DR EMBL; BC054539; AAH54539.1; -, mRNA.  
DR GO; GO:0006897; P:endocytosis; IMP.  
DR InterPro; IPR004765; NP\_C type.  
DR InterPro; IPR003392; Patched.  
DR InterPro; IPR000731; SSD 5TM.  
DR Pfam; PF02460; Patched; 1.  
DR TIGRFam; TIGR00917; 2A060601; 1.  
DR PROSITE; PS50156; SSD; 1.  
SQ SEQUENCE 1277 AA; 142882 MW; 3B42230AACB564E CRC64;

Query Match 34.7%; Score 2400; DB 2; Length 1277;  
Best Local Similarity 40.4%; Pred. No. 8, 7e-16;  
Matches 528; Conservative 233; Mismatches 443; Indels 104; Gaps 27;

QY 11 LMAILLRLAASEPYTTHIQPGVCAFYDECGKRPBELSGSLMTLSNVCLSNTPARKITGDH 70  
DB 8 LGLLLLLCPAQVFSQ-----SCVWYGRCG--IATGD--KRYNCKYSGPKPLPKDG 55  
QY 71 LILQKICPRLYTGNTQAACSAKQVLSLEASLITKALTRCPACDNFVLHGNTCS 130  
DB 56 YDLVQELCRGLFP-DNVSICDDIQQLQTLKSLQPLQPLSRCSCFYNTMLTFCBLTCS 114  
QY 131 PNOSLFINTVTRVAQAGQLPA---VYAEAFYQHSFAEOSYDSCSRVVPAAATLAVG 186  
DB 115 PQSOFILANTATEDYDFDPTQENKTNYKELEYFGQSFANAMYNACRDVEAPSSNEKALG 174  
QY 187 TWCQVYGSALCNAGKWLNPQDGTNGLAPLDI-----TFHLLPEQAGVSGIQLNEGVA 241  
DB 175 LILCGDAPRA-CNATWIEWMFKNDGQAPFTIIPVSDLSII-----GMEPRNATK 225  
QY 242 RCNESQGDVATCSCQDCAASC-----PAIARPOLDSTF-----YLGOMHG 283  
DB 226 GCNESVDVYTGSCQDCSIYCGPKQPPPPMPRIRIGLDMYIMVTVYAFLFVFFG 285  
QY 284 SILVLIILCSYFAVVTILLVGFRRVAPARDKSRMDPKKGTSLSDYLSSTHTLLGQFPQG 343  
DB 286 ALLAWCHRRRRYFSEYTPIDSNIFSVNNS-----DKESASGCDLGAADFCLARMFTK 341  
QY 344 WGTWASMPPLTILVLSVFPVALAGLVFELTTPDPLVLSRPNQASERAFHGHQFP 403  
DB 342 WQAFVCRNPPTCIIFFSLAFITVCGSGLVFVOVITNPVLSAPHSQALKEVEYFKHGP 401  
QY 404 PFRTOQVILTAANRSSRYDSILSLPK-NFSGILDLDELLEBLEQERLRLQVWSPAQ 462  
DB 402 PFRTOQVILTAANRSSRYDSILSLPK-NFSGILDLDELLEBLEQERLRLQVWSPAQ 462  
QY 463 RN---ISLQDICYAPLNDPNTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQV--- 516  
DB 457 YNNEVTYLODLCVADLSPYK---NCTIMSVLNTYQNSHAVL-----DSQVDD 502  
QY 517 -----DMKDHPLCYCANAPLTPFDGTALALSCADYGAAYVFPPLAGYKGDYSABALI 571  
DB 503 FYIYADVHTHFYICRAPASLNDTSLHLGPCICGTGGVFPFVLVAGVDDQYNNATLV 562  
QY 572 MTFSLNTPACGPRLQAQKLEAEFLERARAFORMAAGFOVTFPAERSLEDEIRRTAE 631  
DB 563 ITPFNNTYNTDERQRQAMAEKEKISFYQYK---ENLTISFAERSIDELNRENS 619  
QY 632 DIPAFATSYIYIFLISLALSSYSSWSRWYDSKATLGLGVAVVAGVMAAMGEFFSYLG 691  
DB 620 DVFTYIISYVWFLYISLALGHIGCSRLVDSKISLIGAILVLSVACISGIFSYWG 679  
QY 692 TRSSVLIVQVPELVLSGADNIFIFVUEYQRLPRRPERPERVHIGRLARVAPBMLCS 751  
DB 680 MPLTLIVLEVPFLVAVGVNIFLVQTYORDELOETLDQOGRILGAVAPMPFSS 739  
QY 752 LSEALCFGLATPAPVATFALTSGLAVILDFLQMSAFVALLSLDSKROBASRLDYVC 811  
DB 740 PSETAFAFFGALLSPAVHTSLPAGMAVLIDFLQITCFVSLGLDILKROKHLIDILC 799

```

OY      812  CVKQOELPBPPOG-----BGLLGFQOKAVAPELLHMTIRGVLLFLALPEVSLYSXCH1  867
Db      800  CVRGAD---DQGSASHASESYLPRFEKKNFAPULLKDWLRPIVAAFFGVLSFYAAVYNNK  856
OY      868  SVGLDOELALPKDSYTLDDYFLFLNNKPYFVGAPVYFVYTLGYNFSEAGMNAICSSAGCN  927
Db      857  DIGLDQSLSMNDSYVIDYFQSLAQYHSGPPYFVLEEGNNYSRKQNNVCGMGCDN  916
OY      928  FSFTQKIOYATEFPBQSYLAIIPASSWVDFIDMLTP--SSCCRLYISGCKDKFCFSPSYN  986
Db      917  DSLVOQIINAAELDPTVYRVGAPSSMIDYFDWVSPQSSCCRLY---NVTHQFCNASYMD  973
OY      987  LNCCKNGMSIT--MGSVRPSVEQFHKYLLDMPFLNDRNICCPYGGJLAAYSTYNLSDGVL  1045
Db      974  PTCV--CRPLTPBEGQRPQCKGEFMKFLPMFLSDNPNPKCGKGAHYASANYIGDDTYI  1032
OY      1046  -ASREMAVHKPLKNSQDYTEALRAARELANITADLRKVPSTPAFEVFPYTTINVEYEQ  11040
Db      1033  GATYFMYTHLTIKTSADYTDAMKKARLTIASNTETMBR--KQSD--YRVFSPSYFVFEYEQ  10899
OY      1105  YLTIIIPBGLFMLSCLVPEFPAVSCLLLDGDRSGLLNLSTYMIIVDYVGMALMDISYN  11648
Db      1090  YLTIIDDTIFMLNLSGSIPLVTLVAVLCELMASAVIMCITIAMILVNNFQWMLMGISLN  11499
OY      1165  AVSLINLVSAGVMSVEFVSHITRSPAISTKPTMTLERAKEATISMSGAFAVAMTNTJFI  12244
Db      1150  AVSLVNLVMSGCIStVEFCSHITRAETMTKGSRVARAEALAHMSSSVFSGITLTLPFGGI  12099
OY      1225  LVYGLAKAQOLIOIFFRLNLTLTGLLHGLVFLFVLIISTYGPDPVNP  1272
Db      1210  VLVAFKSOIPEIIFYFRMYLAMLGATHGLFLFVLLSTYGPSPNKA  1257

```

```

12 RESULT
13 Q9N000_FELCA PRELIMINARY; PRT; 1276 AA.
14 ID Q9N000;
15 AC Q9N000;
16 DT 01-OCT-2000 (TREMBLrel. 15, Created)
17 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
18 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
19 DE Nlemaan-Pick type C1 disease protein.
20 GN Name=NPC1;
21 OS Fells silvestris catus (Cat.).
22 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
23 OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;
24 OC Felis.
25 OX NCBI_TaxID=9685;
26 RN [1]
27 RP NUCLEOTIDE SEQUENCE.
28 RC TISSUE=Kidney;
29 RA Murkawa M.; Freeman M.W.;
30 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
31 EMBL; AF258783; AAF72187.1; -; mRNA.
32 DR GO; GO:0016021; C:integral to membrane; IEA.
33 DR GO; GO:0005764; C:lysosome; IEA.
34 DR GO; GO:0016020; C:membrane; IEA.
35 DR GO; GO:0008158; F:hedgehog receptor activity; IEA.
36 DR GO; GO:0030301; P:cholesterol transport; IEA.
37 DR InterPro; IPR004765; NP_C type.
38 DR InterPro; IPR003392; Patched.
39 DR InterPro; IPR000731; SSD_5TM.
40 DR Pfam; PF02460; Patched; 1.
41 DR TIGRFAMs; TIGR00917; 2A060601; 1.
42 DR PROSITE; PSS0156; SSD; 1.
43 SQ SEQUENCE 1276 AA; 141739 MW; CE9B7A0261691256 CRC64;
44
45 Query Match 34.7%; Score 2399; DB 2; length 1276;
46 Best local similarity 40.6%; Pred. No.1e-160;
47 Matches 534; Conservative 222; Mismatches 462; Indels 86; Gaps 29
48
49 7 RGMLLMALILRLAQSPPYTHIQPGYCAFYDECGKNPELSSGLMTLSNVSCISNTSPARKI 66
50 4 RPAALGILILILLC---PAQVLAKS--CIYVSGCG---IASGD-----KRNCKYSGPPKPL 51

```

QY	67	TGHDLLILLOKICPLALYGPMTQA	CCSAKQVLSLPA	SITVAALTTRCPACSDN	VNVN	126						
Db	52	PKDGYDLVQELCEPGEFF	-DNVSLCCDYQOL	LOTLDKNIQLPLQFIS	SRPSCRYNLV	110						
QY	127	NTCSPNOSLFINVTR	-----VAOLGAGOLPA	VAAEAYFQHSFA	QSDSCSRV	182						
Db	111	LTCSPROSQPLNATVATED	YVDVPTNQIKTNKELQYII	IGSSFANAMNAC	GDVAP	170						
QY	183	LAVTMCGVYGSALCA	QWRWLNFGDGTGNGIAP	LDT--PHLEPGA	VSSGIOP	240						
Db	171	KALGILGCKDAEA	-CNATNMIEWF	SKONGCAPPTIPI	FISDLP	224						
QY	241	ARCNESSGDVATSC	CCDCAASC-----PAIAPQ	ALDSFTYLG	QMGSLVLI	291						
Db	225	KGCDESVDEVT	PCSDCSIVCPKQ	PPPPVPMWILG	DAWYVIMWIT	MAFLVFF	284					
QY	292	CSVA-----	VVTILVGRVA--PARDS	KWVDPKGTSL	SPDKLSF	THTLVGQ	339					
Db	285	GAPFLMCTYRKRYV	SETPTIDSI	NA	SVNANDGE-----	ASCCD	LGAABEGCLIR	337				
QY	340	FFQMGWTVASWPL	TLIVLSYIPV	ALAAGLVFTEL	TTDPVELWS	SAPNSQ	ARSEAFHQ	399				
Db	338	LFQMGSPCVAN	PGIIFPSLAF	IAACSSGLVFRV	TTNPVDM	LSAPSQ	ARLEKEY	EDT	397			
QY	400	HFGPFRTNQVIL	TAPNBSRYRYS	LLGPK-ANS	GILDELLELE	GERLRL	QWMS	458				
Db	398	HFGPFRTREQIL	IOAPHTSAHTYOP	YPSGSDVPPG	PLDIALIHQV	DLQTAIEN--	YTA	455				
QY	459	PEAGRNISLD	ICAPLNPDTSLYDC	INSLLQY	PONNRLL	LLTANQ	LMGQSOV--	516				
Db	456	SYNNEVTVLQD	ICVAPLSPYNK--	NCTILSVLNY	FQSHSM	-----	DHETG	DFFYA	507			
QY	517	DMKHFLYCA	NAPLTFKDG	TALALSC	MADYGA	PVFPFLAT	IGYKQDY	SEAEAL	IMTESL	576		
Db	508	DYHHLIXV	APASIND	TLIHP	PCIGTGP	VPFVMTLV	GQDQYNNA	TALVIT	PEV	567		
QY	577	NNYPAGDRL	AOAKLME	EAFL	EEKRAFOR	RMAGH	QVTF	LMERSLE	EDINF	TAEDELP	636	
Db	568	NNYNDTER	LQKAW	MEKEFIN	FKNYKN--	PMLITS	FTTERSTIE	DELNES	NDGIT	FTV	624	
QY	637	ATSYIVIF	PLYLSL	AGSSMS	RMNDKATL	GLGAVV	LGAVMA	ALG	FP	SYGIR	RS	696
Db	625	IISVAMFLY	ISIALGH	IKSC	RLVDSK	ISLGI	ILIV	SVKAS	CS	IGITSY	GI	684
QY	697	VILQVNP	PLVSV	AGDNFI	FVLE	XORLPR	BGE	PRE	HEH	IGR	ALGR	756
Db	665	IIVETIP	PLVAV	AGDNFI	LVQTY	QDER	LAGET	LDQOR	VLGE	VA	PN	744
QY	757	CFEFGAL	PM	PAVTR	FALT	GLA	YLIF	FLLOMS	AF	ALLS	DS	816
Db	745	AFELGAL	SK	MAV	YTFSL	PAG	MAVL	ID	PLD	LOT	ITCV	804
QY	817	ELP	PPGQ	-GEG	LLIG	FQ	KAYAP	FL	LHWIT	RG	VULL	875
Db	805	EDGTS	VQ	AS	ES	CL	FL	FG	HSYS	PL	LD	864
QY	876	AL	PD	SY	LD	FL	PL	N	KY	FEV	GA	935
Db	865	S	MP	D	S	Y	M	D	Y	F	K	923
QY	936	YAT	EP	BO	S	Y	LA	I	P	AS	S	994
Db	924	N	BA	Q	D	S	T	R	I	G	F	979
QY	995	S	I	T	M	-G	S	V	P	S	E	1053
Db	980	P	L	T	G	K	R	P	O	C	G	1038
QY	1053	H	K	P	L	N	S	O	D	T	E	1111
Db	1040	H	T	V	I	O	T	S	A	D	F	1099







DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DE Niemann-Pick C1 protein precursor.  
OS Name=Npcl;  
OC Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
[1]  
RN NUCLEOTIDE SEQUENCE.  
RX MEDLINE=97362324; PubMed=9211850; DOI=10.1126/science.277.5323.232;  
RA Loftus S.K., Morris U.A., Carstea E.D., Gu J.Z., Cummings C.,  
RA Brown A., Ellison J., Ohno K., Rosenfeld M.A., Tagle D.A.,  
RA Pentchev P.G., Pavan W.J.;  
RA "Murine model of Niemann-Pick C disease: mutation in a cholesterol  
RT "homostasis gene.";  
RL Science 277:232-235 (1997).  
RN (2)  
RN SUBCELLULAR LOCATION.  
RX MEDLINE=99145606; PubMed=9990080; DOI=10.1073/pnas.96.4.1657;  
RA Patel S.C., Suresh S., Kumar U., Hu C.Y., Cooney A.,  
RA Blanchette-Mackie E.J., Neufeld E.B., Patel R.C., Brady R.O.,  
RA Patel Y.C., Pentchev P.G., Ong W.-Y.;  
RT "Localization of Niemann-Pick C1 protein in astrocytes: implications  
RT for neuronal degeneration in Niemann-Pick type C disease.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:1657-1662 (1999).  
CC -1- FUNCTION: Involved in the intracellular trafficking of  
cholesterol. May play a role in vesicular trafficking in glia, a  
process that may be crucial for maintaining the structural and  
functional integrity of nerve terminals.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Found in late  
endosomes and lysosomes.  
CC -1- TISSUE SPECIFICITY: Expressed predominantly in perilymphatic  
astrocytic glial processes. Also expressed in heart, spleen, lung,  
liver, skeletal muscle, kidney, testis.  
CC -1- INDUCTION: Activated by the drugs progesterone and U-18666A which  
block cholesterol transport out of lysosomes and by the  
lysosomotropic agent NH4Cl.  
CC -1- DOMAIN: A cysteine-rich N-terminal domain and a C-terminal domain  
containing a di-leucine motif necessary for lysosomal targeting  
are critical for mobilization of cholesterol from lysosomes.  
CC -1- SIMILARITY: Belongs to the patched family.  
CC -1- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use as long as its content is in no way modified and this statement is not  
removed.  
CC -----  
DR EMBL: AF003348; AAB63372.1; -; mRNA.  
DR EMBL: AF003349; AAB63373.1; -; Genomic DNA.  
DR FIR: T30188; T30188.  
DR Ensembl: ENSMUSG0000024413; Mus musculus.  
DR MGI: MGI:1097712; Npcl.  
DR GO: GO:0006897; P: endocytosis; IMP.  
DR InterPro: IPR004765; NP\_C\_type.  
DR InterPro: IPR003392; Patched.  
DR InterPro: IPR000731; SSD\_5TM.  
DR Pfam: PF02460; Patched; 1.  
DR TIGRfam: TIGR00917; 2A060601; 1.  
DR PROSITE: PS50156; SSD; 1.  
KM Glycoprotein; Lysosome; Signal; Transmembrane.  
FT SIGNAL 1 23  
FT CHAIN 24 1278  
FT TRANSMEM 271 291 Niemann-Pick C1 protein.  
FT TRANSMEM 352 372 Potential.  
FT TRANSMEM 623 643 Potential.  
FT TRANSMEM 655 675 Potential.  
FT TRANSMEM 685 705 Potential.  
FT TRANSMEM 761 781 Potential.  
FT TRANSMEM 834 854 Potential.  
FT TRANSMEM 1099 1119 Potential.

FT TRANSMEM 1125 1145 Potential.  
FT TRANSMEM 1196 1216 Potential.  
FT TRANSMEM 1228 1248 Potential.  
FT DOMAIN 621 765  
FT MOTIF 1275 1278  
FT COMPBIAS 250 260  
FT CARBOHYD 71 71  
FT CARBOHYD 123 123  
FT CARBOHYD 138 138  
FT CARBOHYD 166 166  
FT CARBOHYD 223 223  
FT CARBOHYD 229 229  
FT CARBOHYD 415 415  
FT CARBOHYD 460 460  
FT CARBOHYD 479 479  
FT CARBOHYD 525 525  
FT CONFLICT 474 474  
FT CONFLICT 479 479  
SQ SEQUENCE 1278 AA; 142889 MW; 43C7ICE47D283674 CRC64;  
Query Match 34.5%; Score 2385; DB 1; Length 1278;  
Best Local Similarity 40.2%; Pred. No. 1e-159;  
Matches 525; Conservative 233; Mismatches 443; Indels 104; Gaps 27;  
QY 14 LLLRLAQSEPPYTHIQPGYCAFYDECGKNPELSGSLMTLSNVSCLSNTPARKITGDHLIL 73  
12 LLLLCPAAQVQSG-----SCWYGEBCG---IATGD---KRYNCKYSGPPKPLKXGYDL 59  
QY 74 LQICERLRYTGRTYQCCGAKQVLSLEASLSTKALLTRCPACSDNFVNLHCINTGSPNQ 133  
60 VQELCPGLFF-DNVSICCDIQDLQTLKSNLOPLFLSRCPGCFVNLMTLFCGLTSPHQ 118  
QY 134 SLPIINTVRAQAGAGOLPA-----VVAYEAFYHGFSAESQSYDCSRVRVPAATLAVTWC 189  
119 SQPLNTATBEDYFDPEPTKNTNKELEYVGOSFANMNTMNCRDVEASNSNEKALGLLC 178  
QY 190 GYVGSALCAQKWLNFQDGTGGLAPLDI-----TFHLLPQAVSGGIQPLNEGVARN 244  
179 GNDPARA-CNATWIEVMFKNDQGAFTIIPVFSDLSTL-----GMEPMNATKGVN 229  
QY 245 ESGGDVATCSCODCAASC-----PALAPQALDSTP-----YLCMGSLV 266  
230 ESDVETGBCSCODCSIVCGPKRQPPPPMRIMGLDMYVIMVTVYAFVFFGAL 289  
QY 287 LIIILSVFAVYTIILVGRVAPARDKSNVDPKIGTSLSDKLSFSTHTLLGQFPGWGT 346  
290 AVWCHRRRYFSEHYTPIDSNIFSVNS-----DKGASCCDPLGAAPDCLRMFTKMG 345  
QY 347 WVASMPLLTILVSVIPVALAGLVFTLTPDVELMSAPNSQARSEKAFHQHFGPFR 406  
346 FQVRNPCTCIIFFSLATITWCSGSLVQVTTNPVELMSAPHQARLEKEVFDHFGPFR 405  
QY 407 TNOVITLAPNRSSRYKSDLSLLGPK-NFSGILDLLELLELQERLRHLOWSPEQRN- 464  
406 TEOLIIQABNTSVHYIEPPYAGADVFPGLPKKEILHQVLTNQ-----LAISEITASYNN 460  
QY 465 --ISLDICVAPLNPNTSLVDCINSILYFQNNRTLLLTANQTLMGOTSOV----- 516  
461 ETVTLQDICVAPLSPYK--NCTIWSVLTNYSNAVL-----DSQVGDFFYI 506  
QY 517 --DMKDHFLYCANAPLTFPDGTALASCNADYGAPVPEPLAIGYKXKDYSEALIMTF 574  
507 YADYHHTFLYCRAPASLNDTSLHQPCLGTGGPVPVLVGGYDDQYNNATALVITF 566  
Db 575 SLNNYPADGPRLAQAKLEWAPLEEMRAPQRMAGFQVTFPAERSLBEINFTTADLP 634  
567 PVNNYNDTERLQRAAWAKEFISFYKNYK--PWLITISFPAERSIBELNRESNDVFE 623  
QY 635 IPATSYIVFLYISLALGYSMSRMYNDSKATLIGGAAVVLGAYMAAMPSTYSGIRS 694  
624 TVTISTVWVFLYISLALGHIGSCRLVDSKSLTAGLILIVSVAGSLGIFSTYGMPL 683  
QY 695 SLVILQVVPFLVLSVGADNIFLVLEYQRLPRRPGEPREVIIGRALGVAPSKMLCSLSB 754

Db	684	TLIVIEVLPFLVLAVGVNIFILVQTYORDESLQETLDDQGLIGLGEVATMFLSPSE	743
Qy	755	AICEFLGALTMPAVRTEALTSGLAVILDFLLQNSAFVALSLDSKROEASRLDYCCVK	814
Db	744	TSAFPFGLSSMPRVHTSFLFAGNAVILDFLLQITCFVSLIGLDIKRQEKHHDILCCVR	803
Qy	815	POELPPPGQG---EGLLGFPOKAYAPFLLMHTTRGVLLFLFLAFGVSLYSKCHISVG	870
Db	804	GAD--DOGGSHASSESYLFRFFKQVFAILLKMDLPRIVVAVFVGVSFSFAVNVKPDIG	860
Qy	871	LDQELALPKQSYLLDYFLFLNRYPEVGAAPVFTTLTGNFSSSEAGMNAICSSAGCNPSF	930
Db	861	LDQISLSMNDIVYINLFKSLAQYHSHGPPVFLIEEGYNYSRKQNNVCGMGCDNDL	920
Qy	931	TOKIOYATEPEQOSYLAIPASSWDDFTDMLTP--SSCCRLYISGENKDKFCPSTVNSLNC	989
Db	921	VQGFNAAEIDTYTRGVAPSSWIDYFDWVSPQSSCCRLV---NVTHQFCNASVMDPTC	977
Qy	990	LKCKMSIT--MSSVAPSVQEPFKHLYLPMFLNDRPNICPFGGLAAYSTVNLSDQVL--AS	1047
Db	978	VR--CRPLPEPKQKORQGEKFMKFLPMFLSDNPNPCGCGGAAAGSAAVIVGDDTTYGAT	1036
Qy	1048	RFMAYHKELKKSQDYTEALRAARELANITADLRKVPCTDPAFEVFPYITNVFYEQYLT	1107
Db	1037	YFMTHYHTLTKLSADYTDMAKKARLIASITETMRS--KSD--YRFPFISVYVVFYEQYLT	10933
Qy	1108	ILPEGLFMLSCLVETFAVSCLLGLDLRSGLNLNLSTVMILVDTGVFMALMDISYNAS	1167
Db	1094	IIDDTIFMLSISLGISFLVTLVVLGCELSANVIMCITIAMILVNVFGVMTLMSISLNAS	11533
Qy	1168	LINLVASNGMSEVPFSTHTRSPALSTKRTWLERAEALITSMGSANFPAVAMTNLPGILVL	1227
Db	1154	LNLVMSGSGISVEFCFSTHTTRAFMTSTKSRVSRAEALAHNGSSVFSGITLTKGGIVLL	12133
Qy	1228	GLAKAQILQIFFPRLNLITTLGLHGVLPVLISYVGPVNP 1272	
Db	1214	AFALSGQFEIFYFMYTLAMVLLGATHGLIFLPVLLSYIGPSVNKA 1258	
RESULT 15			
09GLC9 BOVIN			
ID	09GLC9	BOVIN PRELIMINARY; PRT; 1277 AA.	
AC	09GLC9		
DT	01-MAR-2001	(TREMBLrel. 16, Created)	
DT	01-MAR-2001	(TREMBLrel. 16, last sequence update)	
DE	01-OCT-2003	(TREMBLrel. 25, last annotation update)	
DT	Niemann-Pick type C1	disease protein.	
GN	Name=NPC1;		
OS	Bos taurus (bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;		
OC	Pecora; Bovidae; Bovinae; Bos.		
NC	NCBI_TaxID=9913;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	TISSUE=Aorta endothelium;		
RA	Mutakawa M., Freeman M.W.;		
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF299073; AAC24620.1; -; mRNA.		
DR	GO; GO:0016021; C:integral to membrane; IEA.		
DR	GO; GO:0005764; C:lysosome; IEA.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0008158; F:hedgehog receptor activity; IEA.		
DR	GO; GO:0030301; P:cholesterol transport; IEA.		
DR	InterPro; IPR004765; NP_C_type.		
DR	InterPro; IPR003392; Patched.		
DR	InterPro; IPR00731; SSD 5TM.		
DR	Pfam; PF02460; Patched; I		
DR	TIGRFAMs; TIGR00917; 2A060601. 1.		
DR	PROSITE; PS0156; SSD; 1.		
SO	SEQUENCE	1277 AA; 14191 MW; 30C91ABFD1AB1CF CRC64;	

Query Match	34.3%	Score 2367.5	DB 2	Length 1277
Best Local Similarity	40.1%	Pred. No. 1.7e-158		
Matches 517	Conservative 227	Mismatch 445	Indels 101	Gaps 26
QY	33	CAFYDECGKPELGSIMLTLSNVSCLSNTAPARKITGHTLILLOKICPRLTYGPNTQACCS	92	
Db	25	CINWGECC---IASGD---KRNCRSGSPPPELPQDGYDLVQELGCGFFFG-NVSLCCD	76	
QY	93	AKQVLVSLEASLITKALLTRCPACSDNFVNLHCNHTCSPNQSLFINVTRVAQL--GAGQ	149	
Db	77	VQOHTLTKNDLQPLQPLQSRSCPCSFVYLVLPCELTCSPROSGFLANTATEDYVDPAQNQ	136	
QY	150	LPAVVAEAAV-QHSFPAEQSYDCSRPRVAAATLAVTMCQCYGSAALCNAQRLNQCQD	208	
Db	137	TKTNVKELOQYVESEFANMAMTNSCRDVEASSNEKALGLLCGBRASA-CNATNMIETWFN	195	
QY	209	TGNGLAELDT--PHLEPGQAVSGIGLOPNEGVARCNESQGDVPAVSCQDCAASCPAI	266	
Db	196	KDNQQAEPFITTPVPSDLPT-----HGHEPMNNAATKGCDESDVEYTGSCQDCCSAVCGPK	250	
QY	267	ARPD-----ALDSTFYLGQMPGSLVLIILCSYFAVAVTLLVG--FRVAPARDK--	313	
Db	251	PQPEPPEVPMKILIGLDMY-----VIMSTYMAELLVFGCAFPAVMCYKRYF	298	
QY	314	-----SKAVDPKKGTSLSKLSFSHTTLTGQFOQMGTVAAHPELTIIVLSV	360	
Db	299	VSEFTPIDGNI PPSINASDKGPTCCDPLCAAEBAHRLRFEWMSGCVAHPGCVAFSV	358	
QY	361	IPVVALAAGVPELTTDPPELMSAPNSQARSEKAFHDFGPFPRNOYILTAPNNSY	420	
Db	359	AFIACSSGLVFQVITDPVDLMSAPQSRARLEKXFDTHFGPFPRTEQLIIRAPHPPH	418	
QY	421	RYDSLILGPK-NFSGILIDLLELLELQERLHLQVMSPEAQRNLSLODICYAPLNDN	479	
Db	419	IYEPYPSGADVPEGPPLAVNIHQVLDLQNAIE--SITASYNNETVLRDICIYAPLSPN	476	
QY	480	TSLYDCCINSLOYPQNNRTLLLTANQTLMGTSQV--DMKHFLYCANAPLTFKQGT	537	
Db	477	---QNCITLISLVNFFQNSHSVL-----DHQVGDFFPYADYHHTFLYVWAAPASLNDTSL	528	
QY	538	LALSQMDVQAPVPEPLAIGGYGKQVQSEAEALIMFSLNNYPAGDRPLQAOLMEBAFL	597	
Db	529	LHDPCLCTFGGAPFPMVLVLDGYDDQNNNAITALTITPVANNYNDEKLGQAQAMEKEFI	588	
QY	598	EEMRAFORMAAGFOVFTAESLDEIRNTTAEDLPIFATSYLVLEFYLYSLALGYSYW	657	
Db	589	NFYQNYN---PMLTISFKAERSTEDLBNESNDVFTVLISGVMLYISIALGHKSC	645	
QY	658	SRVWDSKATLLGGGAVVLGAVMAMGFFSYGIRSSVLLOVPPVLVSGADNIFIF	717	
Db	646	RRLIVDSKILLIGIAGVLIVSPVACSGFFPSYMGSLPTLLIVIEVLPVLVAVGDNIFIL	705	
QY	718	VLEVQRLPRRPGREVRHIGRALGRVAPSMULLCSLSAICFELGALTPMAVPEALTSQ	777	
Db	706	VQYQIDBERIQGELTLDQVGRVILGEVAPSMFLSSFAETVAPFGLGSLVMPAVHFTSLPAG	765	
QY	778	LAVILDELQMSAFVALSLSDSKROEASRLDVCVCVPOELPPG-----QGEGLL	828	
Db	766	MAVILDELQITCTCVSLGIDIKQENQOIDLVCV-----CGAADAAGIOASSSCL	817	
QY	829	LGEFOKAYAEFLHMITRGVLLFLTALFQVSLYSMKHISVGLDQELAPKDSYLLDYFL	888	
Db	818	FRFRNSYAPFLDKDMWRPLVAVVAFGVLSFSLAVLNKVEIGLQDLSMPDDSVATDYFQ	877	
QY	889	FLNRYEEVGAAPVFTVTLGYNFSSSEAGMAICSSAGCNNSFFQKIOYATIEFPQSLAI	948	
Db	878	SLNQYIHAQPPYFVLEEGHDYSTKQNNVCGGLGNDNSLVQOVETPAQLODSYTBIGF	937	
QY	949	PASSWVDDFLDWLTP--SSCRLLYSGNKKQFCPSTYNSLNCJNGCSIT--MGSVRPVE	1006	
Db	938	APSSWIDYFDPMVAPQSSCCRIYNS---TEQFCNAAVYVPTCYR-CAPLTPREGQKROGA	993	
QY	1007	QFHLYLWPLFNDRENICCPKGGILAAVYSTSVNLTSDG--QVLASREMAVHKPLKNSQDYTEA	1065	

```
Db      994  : : | | | | | | | | | | : : | : | | | | : : |
Qy      1066 LRAARELANITADLRKRPCTDPAFEVPEPYTITNVFYOYLITLPEGLFWLSCLVPTFA 1125
Db      1054 : : | : | | | | : : | | | | | | | | | | : : | |
Qy      1126 VSCLLIGLDLRSGILNLSIVMILVDTVGFMALMDISYNAVSLINLVSAGVMSVEFVSHI 1185
Db      1111 VAVVLLGCELMGSAVIMCATIAMILVNMFVGMWLMGISLNAVSLVNLVMSGCISVEFCSHI 1170
Qy      1186 TRSFALSTKPTWLERAKEATISMGSAVPAGVANTNLPGLIVLGLAKAQLIQIFFRLNLL 1245
Db      1171 TRAFVTSTKGSRYERABEALSHMGSSVFSGITLTKFGIIVLAFAKSQIFQIFFRMYLA 1230
Qy      1246 ITLLGLHGLVFLPVILSYVGPDPNPALAL 1275
Db      1231 MVLLGATHGLIFLPVLISYIGPSINKAKSL 1260
```

Search completed: March 22, 2006, 22:55:19  
Job time : 273 secs

**THIS PAGE BLANK (USPTO)**